OPIC
OFFICE DE LA PROPRIÉTÉ INTELLECTUELLE DU CANADA

CIPO
CANADIAN INTELLECTUAL PROPERTY OFFICE

(12) (19) (CA) **Demande-Application**

(21) (A1) **2,219,958**
(22) 1998/01/07
(43) 1999/07/07

(72) LEBERER, Ekkehard, CA
(72) LEEUW, Thomas, DE
(72) WHITEWAY, Malcolm, CA
(72) THOMAS, David Y., CA
(71) NATIONAL RESEARCH COUNCIL OF CANADA, CA

(51) Int.Cl.[6] C12N 15/54, C12N 9/96, C12N 15/62, G01N 33/53, C12Q 1/48, C07K 7/08, C07K 19/00, C07K 14/395, C12N 15/31, C12N 9/12, C07K 14/11

(54) **LE DOMAINE D'INTERACTION DE LA SOUS-UNITE B DE LA PROTEINE G DE LA FAMILLE DES PROTEINES KINAZES STE20P/PAK ET SON UTILISATION DANS LES BIO-ESSAIS**

(54) **THE G-PROTEIN .BETA.SUBUNIT INTERACTION DOMAIN OF STE20P/PAK FAMILY OF PROTEIN KINASES AND USES THEREOF IN BIOASSAYS**

| | | | | | | |
|---|---|---|---|---|---|---|
| Ste20p *(Sc)* | Q03497 | 876 | ean | **S S L** A **P L** V K **L A R** | lkkvaenmdad...//... | 939 |
| Cst20p *(Ca)* | Q92212 | 1209 | ddv | **S S L** S **P L V** K **I A R** | lkkmsesd | 1230 |
| Pak1/Shk1 *(Sp)* | P50527 | 641 | vpv | **S S L** I **P L I** K S **I H** | hsgk | 658 |
| Pak1 *(Hs)* | Q13153 | 525 | kpl | **S S L** T **P L I** A **A A K** | eatknnh | 545 |
| Pak2 *(Hs)* | Q13154 | 505 | kpl | **S S L** T **P L I** M **A A K** | eamksnr | 525 |
| Pak3 *(Hs)* | Q13177 | (473) | kpl | **S S L** T **P L I** M **A A K** | eamksnr | (493) |
| Pak1 *(Rat)* | P35465 | 524 | kpl | **S S L** T **P L I** A **A A K** | eatknnh | 544 |
| Pak2 *(Rat)* | Q62829 | 523 | kpl | **S S L** T **P L I** L **A A K** | eaiknssr | 544 |
| Pak3 *(Rat)* | Q64303 | 507 | kpl | **S S L** T **P L I** L **A A K** | eamksnr | 524 |
| Pak3 *(Rabbit)* | Q29502 | 504 | kpl | **S S L** T **P L I** M **A A K** | eamksnr | 524 |
| Pak3 *(Mouse)* | Q61036 | 523 | kpl | **S S L** T **P L I** I **A A K** | eaiknssr | 544 |
| DPak *(Dm)* | Q24190 | 685 | rpl | A **S L** T **P L I** M **A A K** | eatkgn | 704 |
| Pak1 *(Xen)* | (AF000239) | 504 | kpl | S S **L** T **P** Y **I** I T **G K** | qiakggh | 524 |
| Pak *(Ce)* | (D83215) | 547 | kpl | A **S L** Y Y **L I** V **A A K** | ksiaeasns | 569 |
| MIHCK *(Dd)* | (U67716) | 870 | cns | N G **L** V **P** A **I** M E **A K** | kakeahskfsih | 895 |
| MIHCK *(Ac)* | (U67056) | (279) | gpe | **S** D **L** I **P L V** E R T **K** | neaqrdfsmff | (303) |
| Cla4p *(Sc)* | P48562 | 829 | cdp | K D **L** T S **L L** E W - **K** | e | 842 |
| Cla4p *(Ca)* | (U87996) | 940 | gki | E E **L** A **P L L** E W K **K** | qqqkhqqhkqetsdtgfa | 971 |
| Skm1p *(Sc)* | Q12469 | 643 | csp | E Q **L** K V S **L** K W H | | 655 |
| Pak2/Shk2 *(Sp)* | (U45981) | 570 | cpt | E D **L** K S I **I** F S R **K** | anthin | 589 |

consensus **S S L** $\phi$ **P L** $^{I}{}_{V}$ x $\phi$ $\phi$ $\beta$

(57) The present invention relates generally to signal transduction through G-protein-coupled receptors and more particularly to the interaction between the .beta. subunit of the heterotrimeric G-protein and the Ste20p/PAK family of protein kinases. More particularly, the invention is directed to the identification of the G-protein .beta. subunit interaction domain of Ste20p/PAK family of protein kinases, the Ste20p/PAK interaction domain of G-protein .beta. subunit, to antibodies specific for these interacting domains, the nucleic acid molecules encoding same, to assays, expression vectors, indicator cells, strains, methods and agents which make use of this Ste20p/PAK - G.beta interaction.

## ABSTRACT OF THE DISCLOSURE

The present invention relates generally to signal transduction through G-protein-coupled receptors and more particularly to the interaction between the β subunit of the heterotrimeric G-protein and the Ste20p/PAK family of protein kinases. More particularly, the invention is directed to the identification of the G-protein β subunit interaction domain of Ste20p/PAK family of protein kinases, the Ste20p/PAK interaction domain of G-protein β subunit, to antibodies specific for these interacting domains, the nucleic acid molecules encoding same, to assays, expression vectors, indicator cells, strains, methods and agents which make use of this Ste20p/PAK - $G_\beta$ interaction.

## TITLE OF THE INVENTION

THE G-PROTEIN β SUBUNIT INTERACTION DOMAIN OF STE20P/PAK FAMILY OF PROTEIN KINASES AND USES THEREOF IN BIOASSAYS

## FIELD OF THE INVENTION

The present invention relates generally to signal transduction in cells. More particularly, the present invention relates to signal transduction through G-protein-coupled receptors and especially to the interaction between β subunits of heterotrimeric G-proteins and the Ste20p/PAK family of protein kinases. The invention also relates to assays, expression vectors, strains, methods and agents which make use of this Ste20p/PAK-$G_\beta$ interaction.

## BACKGROUND OF THE INVENTION

The transmission of numerous extracellular signals through the cell membrane, eventually leading to gene expression modulation, is effected through the interplay of G-protein-coupled receptors (GPCR, one of the most ubiquitous transmembrane receptor families) and a heterotrimeric complex of nucleotide-binding regulatory proteins. This complex, also termed tripartite G-proteins or heterotrimeric G-proteins, is comprised of three subunits termed α, β, and γ. These subunits which can transduce the extracellular signal through the GPCR downstream to different signal transduction pathways are the basis for a wide variety of cell signalling functions involved for example in intercellular communication, response to environmental stimuli such as growth factors, hormones, neurotransmitters, physical parameters (such

as light and temperature) and the like. Of importance, the G-protein dependent signalling pathway is conserved in organisms ranging from yeast to man. Due to the structural and functional homologies between the G-proteins in diverse organisms, the yeast *Saccharomyces cerevisiae* is used as a model system for higher eukaryotic cells and organisms. In fact, numerous factors involved in G-protein signalling have been shown to functionally substitute for the yeast equivalents. The tripartite G-protein complex for example, was shown to be functionally reconstituted using mammalian $G_{\alpha}$ and yeast $G_{\beta\gamma}$ (WO 95/21925). In view of the diversity and importance of the signals which induce the G-protein dependent signal transduction pathway, and the importance of the downstream effectors of the G-proteins, the dissection of the interactions taking place in these signal transduction pathways have tremendous fundamental and commercial potential. Furthermore, these interactions represent targets for therapeutic agents. Indeed, the importance of the G-protein-dependent signalling pathway in regulating critical cellular biological functions is demonstrated by the identification of disease conditions which are influenced or determined by mutations in this pathway. For example, the role of GPCRs in disease is reviewed in Coughlin (1994, Curr. Op. Cell. Biol., 6:191-197). Examples of mutations of GPCRs responsible for human diseases have been described (WO 96/41169 and references therein). Moreover, the treatment of a variety of disease conditions is effected through a modulation of the G-protein signalling pathway. For example, agonist analogs of gonadotropin-releasing hormone have been used to treat breast and prostate cancer, endometriosis and non-tumorous ovarian hyperandrogenic syndrome (Pace et al., 1992, Am. Fam. Physician,

44:1777-1782). In view of the critical role played by G-protein signal transduction in cellular homeostasis and disease conditions there remains a need to identify modulators of the G-protein signalling pathways downstream from GPCRs.

The p21-activated protein kinase (PAK) family is a large growing family of regulatory enzymes involved in varied cellular processes ranging from cellular morphogenesis, stress response and apoptosis. The PAK family or Ste20p/PAK family was originally identified based on the property of its kinases to bind to the activated Rho-type p21GTPases Cdc42 and its related protein Rac1. The signature for this family of kinases is a characteristic sequence in the subdomain VIII of the kinase domain (Figure A; Sells et al., 1997, Trends Cell. Biol., 7:162-167).

The Ste20p/PAK family of protein kinases is divided into three groups or sub-families: (1) the so-called true PAKs which contain an N-terminal p21 binding domain (PBD); (2) the pleckstrin-homology (PH) PAKs which also contain a PH-domain upstream of the PBD; and (3) the GCK sub-family exemplified by the germinal center kinase (GCK), which have a long C-terminal region and lacking a recognizable PBD (Figure A).

Like Raf, PAKs link GTPases to a protein kinase cascade. However, unlike Raf, for which the activation by Ras can be attributed in large part to a relocalization of the kinase to the plasma membrane, PAK-p21 interaction alone is sufficient for *in vitro* activation. PAK-Rac and Raf-Ras interactions therefore display both common and different characteristics.

Ste20p kinase, the founding member of the Ste20p/PAK family, shares sequence similarity to protein kinase C, and is required to transmit the pheromone signal from $G_{\beta\gamma}$ to downstream components of the signalling pathway (Leberer et al., 1992, EMBO J., 11:4815-4824). Ste20p/PAK has been shown to be a pivotal point between the G-protein-coupled receptors/G-proteins and the mitogen activated protein kinase (MAP kinase) pathway (Leberer et al., 1997, Curr. Opinion. Genet. & Devel., 7:59-66).

The implication of Ste20p in the activation of a protein kinase cascade prompted the analysis of a similar phenomenon in mammalian cells. Although a definite role for Ste20p/PAKs as major effectors in the stress activated protein kinase cascades (SAPK) has yet to be formally demonstrated, their implication therein has been described (Sells et al., 1997, *supra*). Indeed, the yeast Ste20p regulated pathways such as mating and filamentous growth share similarities with the JNK/SAPK pathway in mammalian cells which is thought to be activated, at least in part, by a cascade of small G-proteins and homologs of Ste20p (Leberer et al., 1997, *supra*). As with Ste20p in yeast, PAKs appear to be involved in morphological responses such as membrane ruffling and the formation of focal adhesions which might be functionally equivalent to mating protrusions in yeast (Leberer et al., 1997, *supra*). Further, the similarity of Ste20p to mammalian p65 PAK (Leberer et al., 1992, *supra* and USP 5,605,825) and of Cdc42p to the mammalian Rho-like guanosine triphosphate Rac1, Cdc42Hs and RhoA, which are known to participate in the activation of the JNK/SAPK signalling cascade and the regulation of actin reorganization in response to extracellular signals, indicates that signal transduction through Ste20p/PAK may be relevant

to the understanding of similar signalling mechanisms in organisms ranging from yeast to mammalian cells (Leeuw et al., 1995, Science, 270:1210-1213). The answers obtained using Ste20p in yeast are therefore of importance in the global understanding of Ste20p/PAK implications in various signalling cascades in eukaryotes in general.

Recent examples have shown the importance of the G-protein-coupled receptor-tripartite G-proteins - Ste20p/PAK interactions (Knaus et al., 1995, Science, 269:221-223; Teo et al., 1995, J. Biol. Chem., 270:26690-26697). It has been established that G-protein coupled receptors can regulate PAKs in mammalian cells. Chemoattractants were shown to rapidly stimulate two human PAKs through the activation of heterotrimeric G-proteins leading to the phosphorylation of $p47^{phox}$, suggesting an implication of G-protein-PAKs in NADPH oxidase regulation, and hence, in inflammatory response of human phagocytic leucocytes. Further, thrombin, which binds to a classical G-protein coupled receptor was shown to activate γ-PAK, a platelet protein kinase displaying significant identity to human p65 PAK, suggesting that PAK may be a part of the thrombin-response signalling complex and platelet function (Teo et al., 1995, *supra*).

Like PAKs, a number of GCK-like PAK members (referred as group (3) above) activate kinase cascades such as the aforementioned Jun N-terminal kinase (JNK) cascade, the stress activated protein kinase (SAPK cascade) and the mitogen activated protein kinase (MAPK cascade). Although sequence similarities between GCK and PAK family members seem limited primarily to the kinase domain, the identification of the p21 binding motifs in the rat homolog of

GCK, raises the possibility that other GCK-PAK-subfamily members might have non-recognized PBDs (Sells et al., 1997, *supra*).

The recent identification of HIV's essential protein Nef as associating with and activating at least one PAK-like kinase further indicates that PAKs and homologs thereof have the potential to play an important role in animal diseases and in human diseases in particular (Sells et al., 1997, *supra*).

The mating-pheromone response in yeast provides a genetically tractable system to study structure/function relationships of the G-protein-Ste20p signal transduction pathway and related pathways *in vivo*. In view of the high degree of functional and structural homologies between the G-proteins and downstream effectors such as the Ste20p/PAK proteins, the yeast system has the potential to provide critical insights into signal transduction pathways in higher eukaryotes (Leberer et al., 1992, EMBO J., 11:4805-4813).

The yeast mating-response MAP kinase cascade consists of Ste11p (a MAP or extracellular signal regulated kinase kinase (MEK) kinase homolog), Ste7p (a MEK homolog) and the partially redundant MAP kinase homologs Fus3p and Kss1p (Leberer et al., 1997, *supra*). Activation of this cascade through binding of pheromones to G-protein coupled receptors induces cellular processes which are typical of differentiating cells, including growth arrest in $G_1$ of the cell cycle, differential gene expression, and polarized morphogenesis which leads to the formation of mating-specific projections (Leberer et al., 1997, *supra*). $G_\beta$-mediated activation of this cascade involves Ste20p (a MEK kinase kinase) and the MAP kinase scaffolding protein Ste5p (Leberer et al., 1997, *supra*). PAKs, a subgroup of mammalian Ste20p homologs, can

be activated by either the small G-proteins Cdc42 and Rac or by heterotrimeric G-proteins in various signalling pathways (Sells et al., 1997, *supra*). The Cdc42p binding domain of Ste20p has been shown to be dispensable for pheromone signalling in yeast suggesting that activation of Ste20p in response to pheromone occurs in a manner independent of Cdc42p (Peter et al., 1996, EMBO J., 15:7046-7059; Leberer et al., 1997, *supra*).

The importance of the Ste20p/PAK family of protein kinases is supported by the significant functional and structural conservation thereof throughout evolution. The recent discovery that certain GCK/PAK subfamily members may also couple with GTPases raises the possibility that PAKs in general may mediate GTPase functions. In view of the critical and often essential roles of such Ste20p/PAK interactions in fundamental and diverse cellular processes, and the conservation of the structure/function relationship of PAKs throughout evolution, there is a tremendous need in dissecting and understanding the molecular determinants involved in Ste20p/PAK-G-protein interactions. Such dissections and understandings might shed a light on the possibility that differential regulation by heterotrimeric and small G-proteins may contribute to Ste20p/PAK specificity on the downstream MAP kinase module, and may explain how the same protein kinase module may regulate different developmental pathways within the same cell.

The present invention seeks to meet these and other needs.

The present description refers to a number of documents, the content of which is herein incorporated by reference.

## SUMMARY OF THE INVENTION

The invention concerns the identification of the domains implicated in the Ste20p/PAK- Ste4p/$G_\beta$ interaction. More particularly, the invention relates to the $G_\beta$ interaction domain of Ste20p and homologs thereof.

The present invention relates to the identification of the molecular determinants of Ste4p/$G_\beta$ interaction in Ste20p/PAK. The invention further relates to the identification of a Ste20p/PAK interaction domain in Ste4p/$G_\beta$.

Also, the invention relates to a characterization of the molecular determinant of a Ste20p/PAK interaction domain in Ste4p/$G_\beta$.

The present invention further relates to isolated polypeptides containing a Ste4p/$G_\beta$ interaction domain of Ste20p/PAK.

As well, it relates to isolated polypeptides containing a Ste20p/PAK interaction domain of Ste4p/$G_\beta$.

Further, the invention relates to epitope-binding portions of the polypeptides of the present invention.

In a preferred embodiment, the Ste4p/$G_\beta$ interaction domain of Ste20p/PAK comprises the amino acid sequence as set forth in the consensus sequence SSL$\phi$PLI$_V$X$\phi\phi\beta$ and as set forth in SEQ. ID. NO.: ID. NO.:27. In a particular embodiment, the Ste20p/$G_\beta$ interaction domain of Ste20p/PAK comprises an amino acid sequence in accordance with the above consensus sequence. Examples of such sequences include sequences as set forth in SEQ. ID. NO.: ID. NOs.:1, 2, 4-11 or derivatives or fragments thereof. Ste20p/$G_\beta$ interaction domains having a sequence with significant homology to the consensus are also provided

for example in SEQ. ID. NO.: ID. NO.:3, 12 and 13 or derivatives or fragments thereof.

In another embodiment, a Ste4p/$G_\beta$ interaction domain of Ste20p/PAK comprises a more divergent amino acid sequence as set forth in SEQ. ID. NO.: ID. NOs.:14-20 or derivatives or fragments thereof, as compared to the above-listed consensus sequence.

In yet another preferred embodiment, the Ste20p/PAK interaction domain of Ste4p/$G_\beta$ comprises the amino acid sequence as set forth in SEQ. ID. NO.: ID. NOs.:21-25 or derivatives or fragments thereof.

The invention in addition relates to nucleic acid sequences encoding a Ste4p/$G_\beta$ interaction domain of Ste20p/PAK and to nucleic acid sequences encoding a Ste20p/PAK interaction domain of Ste4p/$G_\beta$. In one particular embodiment, the nucleic acid sequences encoding a Ste4p/$G_\beta$ domain of Ste20p/PAK encode the amino acid sequence as set forth in one of SEQ. ID. NO.: ID. NOs.:1-13 or functional derivatives thereof, in SEQ. ID. NO.: ID. NOs.:14-20 or to a nucleic acid sequence which hybridizes thereto under high stringent conditions or is at least 90 % identical to such nucleic acid sequences encoding the Ste4p/$G_\beta$ binding domain of the present invention.

In another embodiment, the nucleic acid sequence encoding the Ste20p/PAK interaction domain of Ste4p/$G_\beta$ encodes the amino acid sequence as set forth in SEQ. ID. NOs.:21-26 or derivatives or fragments thereof or to a nucleic acid sequence which hybridizes thereto under high stringent conditions or is at least 90 % identical to nucleic acid sequences encoding the Ste20p/PAK interaction domain of the present invention. In a preferred embodiment, the nucleic acid

sequences of the present invention are as set forth in SEQ. ID. NOs.:28 and 30, derivatives or fragments thereof, or nucleic acid sequences which hybridize thereto under stringent conditions or are at least 90% identical thereto.

The present invention also seeks to provide a recombinant nucleic acid molecule comprising an isolated nucleic acid of the present invention operably linked to a promoter element; cells containing same, and vectors and host cells harboring such vectors for expressing the polypeptides of the invention.

The present invention also seeks to provide antibodies directed to the polypeptides or epitope bearing portions thereof as well as to hybridomas producing monoclonal antibodies directed against such polypeptides.

The invention further seeks to provide methods and compositions to screen for compounds having the ability to modulate a signal transduction pathway through their modulation of the Ste20p/PAK - Ste4p/$G_\beta$ interaction. In one aspect of the present invention, the compound inhibits the Ste20p/PAK - Ste4p/$G_\beta$ interaction and uncouples the G-protein receptor from downstream cascades. In another aspect, the agent enhances the Ste20p/PAK - Ste4p/$G_\beta$ interaction, thereby inducing the activation of a downstream signal transduction cascade. In a particular aspect of the present invention, the abilities of a compound(s) to modulate a signal transduction pathway through their modulation of the Ste20p/PAK - Ste4p/$G_\beta$ interaction is assessed by measuring effects on cellular metabolism. In a particular embodiment, this assessment is made through the use of yeast cells as indicator cells and the effect of the test compound(s) observed through the mating ability of the yeast cells. In

another particular embodiment, this assessment is made through *in vitro* means well known to the person of ordinary skill. Non limiting examples of such *in vitro* means include enzyme-linked immunosorbent assays (ELISA) or other immunological assays, filter binding assays, scintillation proximity assays and the like. Once identified such Ste20p/PAK - Ste4p/$G_\beta$ modulating agents can be used as lead compounds to search for drugs, that can modulate a particular signal transduction pathway.

The present invention is also directed to pharmaceutical compositions for controlling diseases which are dependent on the interaction between Ste20p/PAK and Ste4p/$G_\beta$. As well, the invention relates to the administration of such compositions to an animal suffering from a disease which is dependent on the aforementioned interaction.

Accordingly, the present invention also seeks to provide an assay kit for screening and identifying compounds which modulate the Ste20p/PAK - Ste4p/$G_\beta$ interaction wherein the kit contains a first polypeptide comprising a Ste4p/$G_\beta$ interaction domain of Ste20p/PAK and a second polypeptide comprising a Ste20p/PAK interaction domain of Ste4p/$G_\beta$, and wherein the interaction of the interacting domains is assayable.

The present invention in addition seeks to provide a method for screening and identifying compounds which modulate the Ste20p/PAK - Ste4p/$G_\beta$ interaction, comprising the step of incubating a compound in admixture with a substantially purified first and second polypeptide, wherein the first polypeptide comprises a Ste4p/$G_\beta$ interaction domain of Ste20p/PAK and the second polypeptide comprises a Ste20p/PAK interaction domain of Ste4p/$G_\beta$, and determining the extent to which the compound modulates the interaction between the two

polypeptides as compared to a control incubation in the absence of the compound.

In a particular aspect, the present invention seeks to provide a method of controlling diseases, dependent on an interaction of Ste20p/PAK and Ste4p/$G_\beta$ in an animal such as a mammal and to pharmaceutical compositions therefor.

In addition, the present invention seeks to provide a non-human organism containing the nucleic acid molecule encoding an interaction domain of the present invention. The present invention also seeks to provide a non-human organism containing a knock-out of an interaction domain of the present invention.

The polypeptides and nucleic acid sequences of the present invention have utility in designing *in vitro* and *in vivo* experimental models. Such experimental models enable the screening of large collections of synthetic, semi-synthetic, or natural compounds for therapeutic use in Ste20p/PAK - Ste4p/$G_\beta$-dependent diseases or applications. The present invention also enables the identification of signalling pathways converging at the Ste20p/PAK - $G_\beta$/Ste4p interaction.

The applicant is the first to demonstrate a direct interaction between Ste20p/PAK and $G_\beta$/Ste4p. Before the present invention, it was not clear whether Ste20p PAK and $G_\beta$ interacted. In addition, the applicant is the first to identify the domains involved in the interaction of Ste20p/PAK with Ste4p/$G_\beta$, of relevance to the understanding of signal transduction in all eukaryotic organisms.

In accordance with the present invention, there is therefore provided polypeptidic regions involved in the interaction of Ste20p/PAK and Ste4p/$G_\beta$. As well there is provided nucleic acid

molecules encoding such interacting domains. Further, there is provided fusion proteins comprising the interaction domains of the present invention, nucleic acid molecules encoding same and cells harboring those nucleic acid molecules.

In accordance with the present invention, there is also provided, assays and methods for the identification of compounds which modulate the Ste20p/PAK - Ste4p/$G_\beta$ interaction.

In accordance with the present invention, there is additionally provided methods of treatment and uses of compounds which modulate Ste20p/PAK - Ste4p/$G_\beta$ interaction as well as pharmaceutical compositions containing same.

It shall also be understood, that since there is significant homology between the different members of the Ste20p/PAK family members and between the evolutionary divergent Ste4p/$G_\beta$ sequences (see below), that the person of ordinary skill, will be able to adapt the teachings of the present invention in a variety of ways, with amino acid and nucleic acid sequences from different animals and organisms.

## BRIEF DESCRIPTION OF THE DRAWINGS

Having thus generally described the invention, reference will now be made to the accompanying drawings, showing by way of illustration a preferred embodiment thereof, and in which:

Figure A (prior art) shows a structural comparison of the extended p21-activated Ste20p/PAK family of protein kinases (Sells et al., 1997, Trends in Cell Biol., 7:162-167);

Fig. 1 shows the association of Ste4p with Ste20p and Ste5p in yeast cells. (A) Time course of pheromone-induced Ste20p

binding to HA-Ste4p. HA-Ste4p expressed from the *STE4* promoter in cells deleted for *STE4* was immunoprecipitated after treatment with 1 μM α-factor. Relative amounts of Ste20p and HA-Ste4p were determined in Western blots (*see example in upper and middle panels*) and quantified densitometrically (mean values ±SD, n=3) (*lower panel*). (B) Association of Ste5p with HA-Ste4p expressed from the *STE4* promoter in cells deleted for *STE4* . HA-Ste4p immunoprecipitates from exponentially growing (-) and pheromone-treated (90 minutes) (+) yeast cells were analyzed with antibodies to Ste5p (*upper panel*) or HA-Ste4p (*lower panel*). (C) Overexpression of Ste4p leads to binding to Ste20p. HA-Ste4p was overexpressed from the *GAL1* promoter in cells deleted for *STE20* (lane1) or *STE4* (lanes 2 and 3). HA-Ste4p expression was suppressed in glucose-containing medium in cells deleted for *STE4* (lane4). Immunoprecipitates obtained with antibodies to Ste20p (lanes 1 and 2) or the HA-epitope (lanes 3 and 4) were analyzed for the presence of Ste20p (*upper panel*) and HA-Ste4p (*lower panel*). (D) Coimmunoprecipitation of HA-Ste4p and Ste20p truncation mutants. HA-Ste4p and Ste20p$^{495\text{-}888}$ (lanes 1 and 3) or Ste20p$^{495\text{-}877}$ (lanes 2 and 4) truncation mutants were overexpressed from the *GAL1* promoter in cells deleted for *STE20* . HA-Ste4p (lanes 1 and 2) and Ste20p (lanes 3 and 4) immunoprecipitates were analyzed for the presence of Ste20p (*upper panel*) and HA-Ste4p (*lower panel*) by Western blot analyses. Multiple bands of HA-Ste4p and Ste20p represent phosphorylated forms as indicated by phosphatase treatment (data not shown).

Fig. 2 shows the *In vitro*-$G_\beta$ binding assays. (A) Ste4p binds to a sequence carboxyl-terminal to the kinase domain of Ste20p. GST fusions with full length Ste20p (Leberer et al., 1997, *supra*)

(GST-Ste20p FL ) and the indicated Ste20p fragments were incubated with *in vitro*-translated $^{35}$S-Ste4p in the presence (*left, right and* (+) *in middle panels*) or in the absence ((-), *middle panel*) of *in vitro*-translated HA-Ste18p. GST fusion proteins were detected by Western blot analyses with antibodies to GST (*upper panels*). $^{35}$S-Ste4p was detected by autoradiography (*lower panels*). The presence of HA-Ste18p was confirmed by Western blot analyses (data not shown). (B) Summary of the interactions between Ste20p fragments and Ste4p. The interactions were determined by either *in vitro* binding assays[(a)] or coimmunoprecipitations from yeast extracts[(b)]. Conserved residues are underlined in multiple alignments of carboxyl-terminal sequences of Ste20p (Leberer et al., 1992, *supra*), mouse mPAK3 (Bagrodia et al., *supra*), rat PAK (Manser et al., 1994, Nature, 367:40-46) and yeast Cla4p (Cvrckova et al., 1995, Genes Dev., 9:1817-1830), and human PAK, CBD, Cdc42p binding domain. (C) Interactions of $^{35}$S-Ste4p with mouse mPAK3 and yeast Cla4p. GST and amino-terminal fusions of GST with Ste20p, mouse mPAK3 and Cla4p were incubated with *in vitro*-translated $^{35}$S-Ste4p in the presence (+) or absence (-) of *in vitro*-translated HA-Ste18p. Analyses of proteins were performed as described in (A). Relative amounts of $^{35}$S-Ste4p were normalized for relative levels of full length GST fusion proteins containing the intact carboxyl-terminal Ste4p binding site. Data are given as percent of the amount of $^{35}$S-Ste4p bound to GST-Ste20p (mean values ±SD, $n \geq 3$).

Fig. 3 shows the mutational analyses of the association of Ste4p with Ste20p. Interaction of Ste4p mutants with Ste20p and Ste5p. Fusions of GST with Ste20p and Ste5p were incubated in the presence of *in vitro*-translated HA-Ste18p with wild-type $^{35}$S-Ste4p or the

indicated dominant-negative $^{35}S$-Ste4p mutants (Leberer et al., 1992, *supra*). Relative amounts of the GST fusion proteins and of $^{35}S$-Ste4p were quantified by densitometric evaluation of Western blots and autoradiographs, respectively. Data are given as percentage of binding of wild-type $^{35}S$-Ste4p (mean values ±SD, $n \geq 3$).

Figure 4 shows a model for the role of Ste20p in the activation of the pheromone response pathway.

Figure 5 shows multiple alignments of the $G_\beta$-binding sequence of Ste20p with the homologous regions of related protein kinases of the Ste20p/PAK family. All accession numbers are from the Swiss Prot and PIR or GeneBank (in parentheses) databases. Numbers to the left of the first residue from each sequence indicates the position of this residue in the protein sequence (where 1 is the initiator Met). Number to the right depicts the position of the carboxyl terminal residue. Numbers in parathesis are from incomplete sequences. The consensus sequence for the $G_\beta$-binding motif is show below (where $\phi$ is either A, I, L, M, S, or T, and $\beta$ is a basic residue). Sc, Saccharomyces cerevisiae; Ca, Candida albicans; Sp, Schizosaccharomyces pombe; Hs, Homo sapiens; Dm, Drosophila melanogaster; Xen, Xenopus; Ce, Caenorabditis elegans; Dd, Dictyostelium discoidium; Ac, Acantamoeba.

Figure 6 shows multiple alignments of yeast Ste4p with mammalian $G_\beta$ subunits (Hgbb1, human Gβ1; Hggb2, human Gβ2; Hggb3, human Gβ3; Mgbb4, mouse Gβ4; Mgbb5, mouse Gβ5). The numbers in parentheses are the Swiss Prot accession numbers.

Other objects, advantages and features of the present invention will become more apparent upon reading of the following non-restrictive description of preferred embodiments with reference to the

accompanying drawings which are exemplary and should not be interpreted as limiting the scope of the present invention.

## DESCRIPTION OF THE PREFERRED EMBODIMENT

The domains involved in the Ste20p/PAK - Ste4p/$G_\beta$ interaction are herein provided. Isolated polynucleotides and oligonucleotides encoding the Ste20p/PAK - Ste4p/$G_\beta$ interaction domains are provided by the present invention. Isolated proteins encoded by these polynucleotides and oligonucleotides are also provided. Examples of amino acid sequences in accordance with the present invention include SEQ. ID. Nos.:1-27, 29, and 31. Examples of nucleic acid sequences in accordance with the present invention and from which fragments and derivatives thereof can be obtained include SEQ. ID. Nos.:28 and 30.

Certain aspects of the present invention also include nucleic acid sequences which are homologous to the nucleic acid sequences of the present invention.

In another embodiment of the invention, the amino acid sequences of the present invention provide sequences for obtaining polyclonal or monoclonal antibodies, chimeric antibodies, humanized antibodies and the like which are specific for the Ste20p/PAK - Ste4p/$G_\beta$ interaction domain.

Alternatively, in another embodiment the present invention provides a simple, rapid high-throughput functional bioassay for identifying compounds that modulate the Ste20p/PAK - Ste4p/$G_\beta$ interaction. These compounds can act either as agonists or antagonists of Ste20p/PAK - Ste4p/$G_\beta$ interaction and signalling functions. In one

embodiment, the assay is an *"in vivo"* experimental model based on the incubation of indicator cells with test compounds and the identification of the test compound as agonist or antagonist of Ste20p/PAK - Ste4p/$G_\beta$ interaction. Alternatively, it is based on the use of an *"in vitro"* experimental model such as an enzymatic assay, binding assay and the like (i.e. examples 8 and 9). Compounds can be tested individually or in pools or libraries. The term "antagonist" refers to a compound which inhibits the interaction between Ste20p/PAK and Ste4p/$G_\beta$, thereby uncoupling signal transduction through G-proteins. Alternatively, the term "agonist" refers to a compound that stimulates such a signal transduction by promoting Ste20p/PAK - Ste4p/$G_\beta$ interaction. The term "modulator" is used herein to refer to a compound or a mixture or pool thereof which positively or negatively affect the Ste20p/PAK - Ste4p/$G_\beta$ interaction.

As used herein, the terms "interaction domains" and "binding domains" are used interchangeably.

As used herein the recitation "indicator cells" refers to cells that express an interaction domain of a Ste4p/$G_\beta$ - Ste20p/PAK and a Ste20p/PAK interaction domain of Ste4p/$G_\beta$, and wherein an interaction between these domains is coupled to an identifiable or selectable phenotype or characteristic such that it provides an assessment of the interaction between the domains. Such indicator cells can be used in the screening assays of the present invention. In a preferred embodiment, the indicator cells have been engineered so as to replace at least one of the endogenous Ste20p/PAK and Ste4p/$G_\beta$ interacting domains of Ste4p/$G_\beta$ and Ste20p/PAK respectively, by a chosen derivative, fragment, homolog, or mutant thereof. Alternatively, the indicator cells are engineered so as to inhibit the expression of at least one of the

aforementioned endogenous interacting domains. The cells can be yeast cells or higher eukaryotic cells such as mammalian cells (WO 96/41169). Preferably, the indicator cells are yeast cells. Non-limiting examples of such cells and vectors are exemplified herein below (i.e. examples 7 and 11). In one particular embodiment, an indicator cell of the present invention which is wild type with respect to mating can be used to test a compound or a library thereof in order to identify same which affect mating. In another embodiment, the indicator cell can be a yeast cell harboring vectors enabling the use of the two hybrid system technology as well known in the art (Ausubel et al. 1994, *supra*). In one embodiment, a reporter gene encoding selectable marker can be operably linked to a control element such that expression of the selectable marker is dependent on the interaction of the Ste20p/PAK - Ste4p/$G_\beta$ interacting domains. Such an indicator cell could be used to rapidly screen at high-throughput a vast array of test compounds. In a particular embodiment, the reporter gene is luciferase, β-Gal or green fluorescent protein. It will be understood that the indicator cell, polypeptides and nucleic acids of the present invention can be engineered to be particularly suited for the expression of heterologous Ste20p/PAK and/or Ste4p/$G_\beta$ proteins (WO 95/21925).

As exemplified herein below in one embodiment, at least one of a Ste20p/PAK and Ste4p/$G_\beta$ interaction domain of the present invention may be provided as a fusion protein. The design of constructs therefor and the expression and production of fusion protein are well known in the art (Sambrook et al., 1989, in Molecular Cloning - A Laboratory Manual, Cold Spring Harbor Laboratories, and Ausubel et al., 1994, Current Protocols in Molecular Biology, Wiley, New York). In

certain embodiments, it might be beneficial to introduce a protease cleavage site between the two polypeptide sequences which have been fused. Non-limiting examples of such fusion proteins include a hemagglutinin - Ste4p/$G_\beta$ fusion protein and a Ste20p-GST fusion. In certain embodiments, it might be beneficial to fuse the interaction domains of the present invention to signal peptide sequences enabling a secretion of the fusion protein from the host cell. Signal peptides from diverse organisms are well known in the art. Bacterial OmpA and yeast Suc2 are two non-limiting examples of proteins containing signal sequences.

As used herein, the term "compound" is used broadly to refer to natural, synthetic or semi-synthetic compounds. The term "compound" therefore denotes for examples macromolecules, cell or tissue extracts (from plants or animals). Non-limiting examples of compounds include nucleic acid molecules, peptides, antibodies, carbohydrates and pharmaceutical agents. The agents can be selected and screened by a variety of means including random screening, rational selection and by rational design using for example protein or ligand modelling methods such as computer modelling. The terms "rationally selected" or "rationally designed" are meant to define compounds which have been chosen based on the configuration of the interaction domains of the present invention. As will be understood by the person of ordinary skill, macromolecules having non-naturally occurring modifications are also within the scope of the term "compound". For example, peptidomimetics, well known in the pharmaceutical industry and generally referred to as peptide analogs can be generated by modelling as mentioned above. Similarly, in a preferred embodiment, the polypeptides

of the present invention are modified to enhance their stability. It should be understood that in most cases this modification should not alter the biological activity of the interaction domain. The compounds identified in accordance with the teachings of the present invention have a therapeutic value for the treatment of diseases or conditions which are dependent on Ste20p/PAK - Ste4p/$G_\beta$ interaction. Such diseases or conditions could include proliferative diseases, inflammatory diseases, apoptosis and the like.

As used herein, the term "selectable marker" is used broadly to refer to markers which confer an identifiable trait to the indicator cell. Non-limiting example of selectable markers include markers affecting viability, metabolism, proliferation, morphology and the like.

As used herein, agonists and antagonists of Ste20p/PAK -Ste4p/$G_\beta$ interaction also include potentiators of known compounds with such agonist or antagonist properties. In one embodiment, agonists can be detected by contacting the indicator cell with a compound or mixture or library of compounds for a fixed period of time. The level of gene expression (e.g. the level of luciferase produced) within the treated cells is then determined. The expression level can be compared to that of the reporter gene in the absence of the compound(s). The difference between the levels of gene expression indicates whether the compound(s) of interest agonize the aforementioned interaction. The magnitude of the level of reporter gene product expressed (treated vs. untreated cells) provides a relative indication of the strength of that compound(s) as an agonist. Alternatively, such an indicator cell can be used to identify antagonists.

For example, the test compound or compounds are incubated with the host cell in conjunction with one or more known agonists held at a fixed concentration. An indication and relative strength of the antagonistic properties of the compound(s) can be provided by comparing the level of gene expression in the indicator cell in the presence of the known agonist, in the absence of test compounds vs in the presence thereof.

It shall be understood that the *"in vivo"* experimental model can also be used to carry out an *"in vitro"* assay. For example, cellular extracts from the indicator cells can be prepared and used in one of the aforementioned "*in vitro*" tests (i.e. example 11). Numerous *in vitro* methods to detect and/or quantify the interaction between two interacting polypeptides are known to the person of ordinary skill. For example, antibodies can be used for this purpose. The conditions and the type of assay can be adapted by the person of ordinary skill as a function of the desired type of information required, the format of the assay, the detection method and the type and nature of the antibody used. Non limiting examples of commonly known immunological assays which can be used to assess the interaction between Ste20p/PAK and Ste4p/$G_\beta$ include radioimmunoassays, ELISA, immunofluorescence-type assays and the like. Immunological assays which can be used in the context of the present invention are described for example in Harlow et al., 1988 (in: Antibody - A Laboratory Manual, CSH Laboratories). As well different type of binding assays, for example direct or indirect, or competitive binding assays can be used. Scintillation proximity-type assays are other non limiting examples of assays which can be used to identify compounds which modulate the Ste20p/PAK - Ste4p/$G_\beta$ interaction.

For certainty, as used herein "Ste20p/PAK" and "Ste4p/$G_\beta$" refer herein to members of the Ste20p/PAK family of protein kinases and to homologs of "$G_\beta$", respectively. Thus, any Ste20p/PAK or any Ste4p/$G_\beta$ family member with the proviso that it comprises the interaction domains of the present invention or nucleic acid sequences encoding same can be used to practice the present invention. For certainty, the sequences and polypeptides useful to practice the invention include without being limited thereto mutants, homologs, subtypes, alleles and the like. It shall be understood that generally, the sequences of the present invention should encode a functional (albeit defective) interaction domain. It will be clear to the person of ordinary skill that whether an interaction domain of the present invention, variant, derivative, or fragment thereof retains its function in binding to its partner can be readily determined by using the teachings and assays of the present invention and the general teachings of the art. As exemplified herein below, the interaction domains of the present invention can be modified, for example by *in vitro* mutagenesis, to dissect the structure-function relationship thereof and permit a better design and identification of modulating compounds. However, some derivative or analogs having lost their biological function of interacting with their respective interaction partner (Ste20p/PAK or Ste4p/$G_\beta$) may still find utility, for example for raising antibodies. Such analogs or derivatives could be used for example to raise antibodies to the interaction domains of the present invention. These antibodies could be used for detection or purification purposes. In addition, these antibodies could also act as competitive or non-competitive inhibitor and be found to be modulators of Ste20p/PAK - Ste4p/$G_\beta$ interaction.

A consensus sequence of the Ste4p/$G_\beta$ interaction domain is herein provided. It shall be clear that a 100% identity to this consensus sequence is not necessary to provide functionality to Ste20p/PAK (binding to Ste4p/$G_\beta$) since for example (and as described below), a serine to alanine substitution at the first aa position thereof (DPak; SEQ. ID. NO.: ID. NO.:12) retains the biological function. The same can be said of SEQ. ID. NO.: ID. NO.:3, since Shk1 of *S.Pombe* complements a *Ste20* gene disruption. More divergent amino acid sequences, as exemplified for example by SEQ. ID. NO.: ID. NO.:17 does not bind, however. Thus, more divergent amino acid sequence such as SEQ. ID. NO.: ID. NOs.:14-20 and especially SEQ. ID. NO.: ID. NOs.:17-20 can be used to identify compounds and/or molecular determinants of the sequence which can stimulate the Ste4p/$G_\beta$ interaction.

Nucleotide sequences are presented herein by single strand, in the 5' to 3' direction, from left to right, using the one letter nucleotide symbols as commonly used in the art and in accordance with the recommendations of the IUPAC-IUB Biochemical Nomenclature Commission.

The present description refers to a number of routinely used recombinant DNA (rDNA) technology terms. Nevertheless, definitions of selected examples of such rDNA terms are provided for clarity and consistency.

As used herein, "isolated nucleic acid molecule", refers to a polymer of nucleotides. Non-limiting examples thereof include DNA and RNA molecules purified from their natural environment.

The term "recombinant DNA" as known in the art refers to a DNA molecule resulting from the joining of DNA segments. This is often referred to as genetic engineering.

The term "DNA segment", is used herein, to refer to a DNA molecule comprising a linear stretch or sequence of nucleotides. This sequence when read in accordance with the genetic code, can encode a linear stretch or sequence of amino acids which can be referred to as a polypeptide, protein, protein fragment and the like.

The nucleic acid (i.e. DNA or RNA) for practicing the present invention may be obtained according to well known methods.

A host cell or indicator cell has been "transfected" by exogenous or heterologous DNA (e.g. a DNA construct) when such DNA has been introduced inside the cell. The transfecting DNA may or may not be integrated (covalently linked) into the genome of the cell. In prokaryotes, yeast, and mammalian cells for example, the transfecting DNA may be maintained on an episome such as a plasmid. With respect to eukaryotic cells, a stably transfected cell is one in which the transfecting DNA has become integrated into the genome so that it is inherited by daughter cells upon replication. The stability of the integrated DNA can be demonstrated by the establishment of cell lines or clones comprised of a population of daughter cells containing the transfecting DNA. Transfection methods are well known in the art (Sambrook et al., 1989, *supra*; Ausubel et al., 1994, *supra*).

"Nucleic acid hybridization" refers generally to the hybridization of two single-stranded nucleic acid molecules having complementary base sequences, which under appropriate conditions will form a thermodynamically favored double-stranded structure. Examples

of hybridization conditions can be found in the two laboratory manuals referred above (Sambrook et al., 1989, *supra*, and Ausubel et al., 1994, *supra*) and are commonly known in the art. In the case of a hybridization to a nitrocellulose filter, as for example in the well known Southern blotting procedure, a nitrocellulose filter can be incubated overnight at 65°C with a labeled probe in a solution containing 50% formamide, high salt ( 5 x SSC or 5 x SSPE), 5 x Denhardt's solution, 1% SDS, and 100 µg/ml denatured carrier DNA ( i.e. salmon sperm DNA). The non-specifically binding probe can then be washed off the filter by several washes in 0.2 x SSC/0.1% SDS at a temperature which is selected in view of the desired stringency: room temperature (low stringency), 42°C (moderate stringency) or 65°C (high stringency). The selected temperature is based on the melting temperature (Tm) of the DNA hybrid. Of course, RNA-DNA hybrids can also be formed and detected. In such cases, the conditions of hybridization and washing can be adapted according to well known methods by the person of ordinary skill. Stringent conditions will be preferably used (Sambrook et al.,1989, *supra*).

As used herein, the term "gene" is well known in the art and relates to a nucleic acid sequence defining a single protein or polypeptide. A "structural gene" defines a DNA sequence which is transcribed into RNA and translated into a protein having a specific amino acid sequence thereby giving rise the a specific polypeptide or protein.

A "heterologous" (i.e. a heterologous gene) region of a DNA molecule is a subsegment segment of DNA within a larger segment that is not found in association therewith in nature. The term "heterologous" can be similarly used to define two polypeptidic segments

not joined together in nature. Non-limiting examples of heterologous genes include reporter genes such as luciferase, chloramphenicol acetyl transferase, β-galactosidase, and the like which can be juxtaposed or joined to heterologous control regions or to heterologous polypeptides.

The term "vector" is commonly known in the art and defines a plasmid DNA, phage DNA, viral DNA and the like, which can serve as a DNA vehicle into which DNA of the present invention can be cloned. Numerous types of vectors exist and are well known in the art.

The term "expression" defines the process by which a structural gene is transcribed into mRNA (transcription), the mRNA is then being translated (translation) into one polypeptide (or protein) or more.

The terminology "expression vector" defines a vector or vehicle as described above but designed to enable the expression of an inserted sequence following transformation into a host. The cloned gene (inserted sequence) is usually placed under the control of control element sequences such as promoter sequences. The placing of a cloned gene under such control sequences is often referred to as being operably linked to control elements or sequences.

Expression control sequences will vary depending on whether the vector is designed to express the operably linked gene in a prokaryotic or eukaryotic host or both (shuttle vectors) and can additionally contain transcriptional elements such as enhancer elements, termination sequences, tissue-specificity elements, and/or translational initiation and termination sites.

As used herein, the designation "functional derivative" denotes, in the context of a functional derivative of a sequence whether

an nucleic acid or amino acid sequence, a molecule that retains a biological activity (either function or structural) that is substantially similar to that of the original sequence. This functional derivative or equivalent may be a natural derivatives or may be prepared synthetically. Such derivatives include amino acid sequences having substitutions, deletions, or additions of one or more amino acids, provided that the biological activity of the protein is conserved. The same applies to derivatives of nucleic acid sequences which can have substitutions, deletions, or additions of one or more nucleotides, provided that the biological activity of the sequence is generally maintained. When relating to a protein sequence, the substituting amino acid as chemico-physical properties which are similar to that of the substituted amino acid. The similar chemico-physical properties include, similarities in charge, bulkiness, hydrophobicity, hydrophylicity and the like. The term "functional derivatives" is intended to include "fragments", "segments", "variants", "analogs" or "chemical derivatives" of the subject matter of the present invention.

Thus, the term "variant" refers herein to a protein or nucleic acid molecule which is substantially similar in structure and biological activity to the protein or nucleic acid of the present invention.

The functional derivatives of the present invention can be synthesized chemically or produced through recombinant DNA technology. all these methods are well known in the art. In view of the conservation of the Ste4p/$G_\beta$ binding domain of Ste20p/PAK throughout evolution (see below), it will be apparent to the person of ordinary skill that sequences from different organisms and animals and chimeras

thereof can be used in accordance with the teachings of the present invention.

As used herein, "chemical derivatives" is meant to cover additional chemical moieties not normally part of the subject matter of the invention. Such moieties could affect the physico-chemical characteristic of the derivative (i.e. solubility, absorption, half life and the like, decrease of toxicity). Such moieties are exemplified in Remington's Pharmaceutical Sciences (1980). Methods of coupling these chemical-physical moieties to a polypeptide are well known in the art.

The term "allele" defines an alternative form of a gene which occupies a given locus on a chromosome.

As commonly known, a "mutation" is a detectable change in the genetic material which can be transmitted to a daughter cell. As well known, a mutation can be, for example, a detectable change in one or more deoxyribonucleotide. For example, nucleotides can be added, deleted, substituted for, inverted, or transposed to a new position. Spontaneous mutations and experimentally induced mutations exist. The result of a mutations of nucleic acid molecule is a mutant nucleic acid molecule. A mutant polypeptide can be encoded from this mutant nucleic acid molecule.

As used herein, the term "purified" refers to a molecule having been separated from a cellular component. Thus, for example, a "purified protein" has been purified to a level not found in nature. A "substantially pure" molecule is a molecule that is lacking in all other cellular components.

In general, techniques for preparing antibodies (including monoclonal antibodies and hybridomas) and for detecting antigens using antibodies are well known in the art (Campbell, 1984, In

"Monoclonal Antibody Technology: Laboratory Techniques in Biochemistry and Molecular Biology", Elsevier Science Publisher, Amsterdam, The Netherlands) and in Harlow et al., 1988 (*supra*). The present invention also provides polyclonal, monoclonal antibodies, or humanized versions thereof, chimeric antibodies and the like which inhibit or neutralize their respective interaction domains and/or are specific thereto.

The term "non-human animals" refers to animals having a transgenic interruption or alteration of an endogenous gene encoding an interaction domain of the present invention (knock-out animal) and/or animals having an interruption into the genome in which a transgene (directing the expression of encoding an interaction domain of, or the present invention) has been introduced. Non-limiting examples of such non-human animals include vertebrates such as rodents, non-human primates, amphibians, reptiles and the like. These animals are prepared in accordance with known methods.

The present invention is described in further detail in the following non-limiting examples.

## EXAMPLE 1

### Yeast strains and manipulations

*S. cerevisiae* strains used herein were W303-1A (*MATa ade2 leu2 trp1 ura3 his3 can1*), YEL206 (W303-1A *ste20Δ-3::TRP1*) (Wu et al., 1995, J. Biol. Chem., 270:15984-15992), YEL155 (W303-1A *ste5Δ::TRP1*) and YEL121 (W303-1A *ste4Δ::LEU2*). Mating assays, analysis of mating projection formation, and growth arrest, induction of *FUS1::lacZ* and complementation assays of the growth defect of cells

deleted for both *STE20* and *CLA4* were carried out as described (Leberer et al., 1997, *supra*; Leberer et al., 1993, Genet., 241:241-254).

## EXAMPLE 2

**Construction of plasmids**

To construct pBTL110 carrying *HA-STE4P/GB* under control of the *STE4P* promoter, a fragment from nucleotides -491 to -1 of STE4Pwas amplified by PCR and cloned into pRS313 (Sikorski et al., 1989, Genetics, 122:19-27). The *BamHI* fragment of pL55 (Whiteway et al., 1995, Science, 269:1572-1575) was then subcloned downstream of the *STE4P/GB* promoter.

To create pBTL38 and pBTL65 carrying *STE4P/GB* and *HA-STE18* under control of the T3 RNA polymerase promoter, *STE4P/GB* and *HA-STE18* were amplified by PCR and ligated into pRS316 and pRS313 (Sikorski et al., 1989, *supra*), respectively.

To create pBTL79, pBTL80, pBTL81 and pBTL82 carrying the $STE4^{D62N}$, $STE4^{K55E}$, $STE4^{N157H/S175P}$ and $STE4^{\Delta F177}$ mutants under control of the T7 RNA polymerase promoter, respectively, the *GAL1* promoter was excised from pGAL-STE4P/GB-D62N, pGAL-STE4P/GB-K55E, pGAL-STE4P/GB-N157H/S175P and pGAL-STE4P/GB-ΔF177, respectively (Leberer et al., 1992, *supra*).

To create pDH171 and pDH172 carrying the $Ste20p^{495-877}$ and $Ste20p^{495-888}$ fragments under control of the *GAL1* promoter, respectively, these fragments were amplified by PCR and subcloned into pRS313GAL (Leberer et al., 1992, *supra*).

To create pBTL83, pBTL84, pBTL146 and pBTL147 carrying fusions of GST with the $Ste20p^{876-939}$, $Ste20p^{876-892}$, $Ste20p^{819-875}$

and Ste20p$^{819-892}$ fragments, respectively, fragments were amplified by PCR and subcloned into pGEX-4T-1 (Pharmacia).

To create a fusion of GST with full length Ste5p, the *STE5* coding region was amplified by PCR and ligated into pGEX-4T-3 (Pharmacia) to yield pVL50.

## EXAMPLE 3

### Oligodeoxynucleotide-directed mutagenesis of *STE20*

pBTL151 and pBTL150 carrying the *STE20* mutants *STE20*$^{CLA4}$, in which the sequence encoding amino acids 879 to 887 of *STE20* was replaced by the sequence of *CLA4* encoding amino acids 832 to 840 (Cvrckova et al., 1995, *supra*), and *STE20*$^{S879A/S880A/P883A}$, respectively, under control of the *STE20* promoter, were created by site directed mutagenesis (Kunkel et al., 1987, Methods in Enzymology, 154:367-382). The mutations were confirmed by sequencing. To create pBTL117 and pBTL118 carrying fusions of GST with the fragments from amino acid 819 to 939 of the *STE20*$^{S879A/S880A/P883A}$ and *STE20*$^{CLA4}$ mutants, respectively, these fragments were amplified by PCR and subcloned into pGEX-4T-2 (Pharmacia).

## EXAMPLE 4

### Immunochemical procedures

Immunoprecipitation experiments with specific antibodies to the HA-epitope (12CA5 monoclonal and rabbit polyclonal anti-HA antibodies were from Babco, Richmond), Ste20p (Wu et al., 1995, J. Biol. Chem. 270:15984) and Ste5p (Wall et al., 1995, Cell 83:1047-1058) were performed according to standard procedures as

described (Whiteway et al., 1995, *supra*;Leeuw et al., 1995, *supra*). For the detection of Ste20p fragments, a secondary sheep antibody specific to rabbit immunoglobulin light chains and a tertiary HRP-conjugated donkey antibody to sheep IgG were obtained from The Binding Site, Lim. Immunoprecipitations were confirmed in at least three independent experiments. For quantitation, immunoblots were evaluated by integrating densitometry using an Epson ES 1200-C densitometer and the NIH Image 1.59 software.

## EXAMPLE 5

### *In vitro* $G_\beta$ binding assays

Plasmids were linearized downstream of the termination codons of the respective genes. *In vitro*-transcription was performed by using either T3 or T7 RNA polymerase and $m^7G(5')ppp(5')G$ capped GTP. *In vitro*-translation of the resulting mRNA was carried out with $^{35}S$-labeled methionine using an *in vitro*-translation kit (Promega).

GST fusion proteins were purified on glutathione-Sepharose beads in 20 mM HEPES buffer pH 7.4, containing 100 mM NaCl, 50 mM NaF, 0.5 M Sorbitol, 2 mM EDTA, 1 mM $Na_3VO_4$, 0.1% Triton X-100, 1% BSA (wt/v) and a protease inhibitor cocktail, and washed 5 times in phosphate buffered saline (PBS) (140 mM NaCl, 2.7 mM KCl, 10 mM $Na_2HPO_4$, 1.8 mM $KH_2PO_4$, pH 7.4) by centrifugation at 1.000 *g*. Proteins (5-10 µg) were incubated with 5 µl of the reticulocyte lysate containing the *in vitro*-translated products in 25 µl of PBS for 10 minutes at 30°C.

The beads were then washed three times with PBS, separated by SDS-PAGE and analysed by autoradiography of Western

blots. Results were confirmed in at least three independent experiments. Immunodetection and evaluation of immunoblots and radiographs were then performed as described above. Data obtained for $^{35}S$-Ste4p were corrected for relative concentrations of the respective GST fusion proteins.

## EXAMPLE 6

### *In vivo* association of Ste20p with $G_\beta$ (Ste4p)

Coimmunoprecipitation experiments were performed to analyze the *in vivo*-association of Ste20p with an influenza hemagglutinin (HA)-epitope tagged version of $G_\beta$ (HA-Ste4p). Antibodies to HA-Ste4p precipitated low amounts of Ste20p (Fig. 1A). An approximately 5-fold increase in the interaction between Ste4p and Ste20p was observed already after 3 minutes of pheromone treatment and maintained for up to 15 minutes of stimulation (Fig. 1A). The initial induction of Ste20p/Ste4p complexes is consistent with the time course described for the stimulation of Far1p (a cyclin inhibitor) and the MAP kinases Fus3p and Kss1p (Chang et al., 1992, Mol. Biol. Cell., 3:445-450; Gartner et al., 1992, Genes Dev., 6:1280-1292) and may be required to activate the MAP kinase cascade for the induction of growth arrest and transcriptional activation. Additional formation of complexes after prolonged treatment with pheromone (Fig. 1A) followed a time course concomitant with the formation of mating projections and accumulation of receptors, Ste20p and Ste4p (Leberer et al., 1997, Curr. Op. Genet. & Dev., 7:59-66; Leberer et al., 1997, EMBO J., 16:83-97) in the tips of mating projections, and could be involved in the control of morphological changes that may require Ste20p dependent phosphorylation of myosin-I or activation of the

PKC pathway (Leberer et al., 1997, EMBO J., 16:83-97; Wu et al., 1996, J. Biol. Chem., 271:31787-31790).

When the pheromone response pathway was activated through overexpression of HA-Ste4p, Ste5p also formed a complex with Ste4p (Whiteway et al., 1995, *supra*). This complex was present in cells without an activated pathway when HA-Ste4p was expressed at wild-type levels (Fig. 1B), and the association was not significantly altered after treatment of cells with pheromone (Fig. 1B), suggesting a constitutive interaction between Ste4p and Ste5p. Constitutive activation of the pheromone signaling pathway through overexpression of HA-Ste4p stimulated the association of Ste4p with Ste20p in the absence of pheromone (Fig. 1C). This association required the function of Ste18p, the γ-subunit of the mating response G-protein, but did not require the presence of Ste5p (data not shown).

## EXAMPLE 7

### Identification of the $G_\beta$ interaction domain of Ste20p

Cells overexpressing the Ste20p$^{495-888}$ fragment were normal in their mating functions, whereas cells overexpressing the Ste20p$^{495-877}$ fragment were defective (Table 1). Briefly, Strain YEL206 deleted for *STE20* was transformed with pDH166 (Whiteway et al., 1995, *supra*), pDH171 and pDH172 carrying either wild-type *STE20* (*STE20*$^{WT}$) or the *STE20*$^{495-877}$ and *STE20*$^{495-888}$ mutant alleles, respectively, under control of the *GAL1* promoter. Mating efficiencies represent mean values ± SD (n=3). Mating functions were analyzed as described (Leberer et al., 1997, *supra*; Leberer et al., 1993, *supra*).

Table 1. Effects of carboxy-terminal truncations on signaling functions of Ste20p

| *STE20* allele | Mating efficiencies (%) | FUS1:: *lacZ* expression | | $G_1$ arrest | Shmoo formation |
|---|---|---|---|---|---|
| | | Basal | Induced | | |
| $STE20^{WT}$ | 89.8 ± 15.5 | 0.1 | 221.2 | + | + |
| $STE20^{495-888}$ | 75.2 ± 7.5 | 1.6 | 147.7 | + | + |
| $STE20^{495-877}$ | 0.02 ± 0.015 | < 0.1 | < 0.1 | - | - |

These results suggest that the region from amino acids 877 to 888 carboxyl-terminal to the kinase domain of Ste20p plays an important role in the pheromone response. This region was also required for coimmunoprecipitation with HA-Ste4p (Figure 1D), underlining the physiological importance of the association between Ste4p and Ste20p in pheromone signaling.

## EXAMPLE 8

### *In vitro* association of Ste20p with $G_\beta$ (Ste4p)

[$^{35}$S]methonine-labeled Ste4p ($^{35}$S-Ste4p) and HA-Ste18p were synthesized in an *in vitro*-translation system and analyzed for their ability to bind to fusions of glutathione S-transferase (GST) with wild-type Ste20p and fragments of Ste20p (Fig. 2A). It was found that a fragment carboxyl-terminal to the kinase domain encompassing residues 876 to 892 was necessary and sufficient to bind $^{35}$S-Ste4p (Fig. 2A). The binding did not depend on the presence of HA-Ste18p (Fig. 2A,C), and HA-Ste18p alone was not able to bind Ste20p (data not shown). As summarized in Fig. 2B, these results, together with data obtained in the immunoprecipitation experiments, suggest that the non-catalytic region from amino acids 876 to 888 of Ste20p represents a binding site for $G_\beta$.

## EXAMPLE 9

### The $G_\beta$ interaction domain of Ste20p is functionally conserved in mouse mPAK3 and in Ste20p/PAK members

Consistent with observations that mammalian PAK isoforms can complement the mating defect of yeast cells deleted for *STE20* (Bagrodia et al., 1995, J. Biol. Chem., 270:22731-22737) and that the $G_\beta$binding site is conserved in these kinases (Fig. 2B), mouse mPAK3 (Bagrodia et al., 1995, *supra*) bound $^{35}$S-Ste4p (Fig. 2C).

## EXAMPLE 10

### Identification of the molecular determinants of the $G_\beta$ interaction domain of Ste20p/PAK

Ste20p and its closely related isoform Cla4p share a redundant function that is essential for cellular viability in yeast (Cvrckova et al., 1995, *supra*). Consistent with observations that only high levels of Cla4p after overexpression partly complement the mating defect of yeast cells deleted for *STE20* (data not shown) and the Ste4p binding site of Ste20p is not well conserved in Cla4p (Fig. 2B), only weak binding of Ste4p to Cla4p was observed (Fig. 2C). These results support the view that residues conserved in the Ste4p binding sites of Ste20p and PAK isoforms (Fig. 2B) contribute to the binding of $G_\beta$. These results also provide a weak consensus sequence for $G_\beta$ binding, the sequence of the $G_\beta$ binding domain of Cla4p. This weak $G_\beta$ binding consensus sequence can be used in assays to identify compounds which can stimulate Cla4p-$G_\beta$ interaction. In a particular embodiment, the assay involves the use of agents to identify agonists of the Cla4p-$G_\beta$ interaction that will enable a complementation of the mating defect of yeast cells deleted for

Ste20p. In another embodiment, the physical interaction between Cla4p and $G_\beta$ could be assessed *in vitro* through gel shifts, immunoprecipitation and the like, as well known to the person of ordinary skill and as shown herein. Further, single mutations (or a combination of mutations) in the $G_\beta$-binding domain of Cla4p could identify the minimal primary structure requirements enabling adequate $G_\beta$-Cla4p binding and perhaps the complementation of the Ste20p null mutant strain mentioned above. A very similar approach is exemplified with single mutations and a triple mutations of the conserved residues of Ste20p (see below).

## EXAMPLE 11

### *In vitro* mutagenesis of the $G_\beta$ interaction domain of a Ste20p/PAK member

Single alterations of the conserved residues S879, S880 or P883 to alanine did not affect the *in vivo*-function of Ste20p (data not shown). However, the triple mutant Ste20p$^{S879A/S880A/P883A}$ in which the highly conserved sequence motif SSLxPL was altered to AALxAL, showed strong defects in mating functions (Table 2). Briefly, for GST (control) and fusions of GST with the carboxyl-terminal fragments from amino acids 819 to 939 of wild-type Ste20p (*STE20*$^{WT}$), the Ste20p$^{S879A/S880A/P883A}$ mutant and the Ste20p$^{CLA4}$ mutant (in which the sequence encoding amino acids 879 to 887 of *STE20* was replaced by the sequence of *CLA4* encoding amino acids 832 to 840; Cvrckova et al., 1995, *supra*), respectively, were incubated with *in vitro*-translated $^{35}$S-Ste4p in the presence of *in vitro*-translated HA-Ste18p. Data are given as relative levels of bound $^{35}$S-Ste4p normalized against binding to wild-type Ste20p$^{819-939}$. For cells deleted for *STE20* were transformed with

pRS313 (control), pSTE20-5 carrying wild-type *STE20* (Leberer et al., 1992, *supra*) ($STE20^{WT}$), pBTL150 carrying the $STE20^{S879A/S880A/P883A}$ mutant, and pBTL151 carrying the $STE20^{CLA4}$ mutant. Proteins were expressed under control of the *STE20* promoter. Finally *in vitro*-kinase activities were determined in immune complexes isolated from YEL206 cells expressing the indicated *STE20* alleles or the inactive $STE20^{K649R}$ mutant (Wu et al., 1995, *supra*) as a control. In vitro-kinase assays were performed as described (Wu et al., 1995, *supra*) with myelin basic protein (MBP) as substrate. Data are given as percentage of MBP phosphorylation by wild-type Ste20p. Similar defects were also observed for the mutant $Ste20p^{CLA4}$, in which the Ste4p binding site of Ste20p was replaced by the equivalent region of Cla4p (Table 2 and Fig. 2B). No differences were found for the *in vitro*-kinase activities of these mutants when compared with wild-type Ste20p (Table 2), and the mutants were found to complement the growth defect of cells deleted for both *STE20* and *CLA4* (data not shown). However, binding to $^{35}$S-Ste4p was strongly reduced when fusions of GST with both mutant versions were analyzed in the *in vitro*-binding assay (Table 2). Thus, the mating defects of these Ste20p mutants correlated with their reduced ability to bind Ste4p.

Table 2. Effect of carboxy-terminal mutations in Ste20p on *in vitro*-binding to Ste4p, *in vivo*-signaling functions and *in vitro*-kinase activities

| *STE20* allele | Binding to $^{35}$S-Ste4p (%)[a] | Mating efficiencies (%)[b] | FUS1:: *lacZ* expression[b] | $G_1$ arrest[b] | Shmoo formation[b] | Kinase activity (%)[c] |
|---|---|---|---|---|---|---|
| *STE20*$^{WT}$ | 100 | 72.7 ± 6.8 | 274 ± 41 | + | + | 100 |
| *STE20*$^{S879A/S880A/P883A}$ | 15.2 ± 7.8 | 0.039 ± 0.002 | 14 ± 1.8 | - | - | 104 ± 11.2 |
| *STE20*$^{CLA4}$ | 11.1 ± 9.3 | 0.016 ± 0.001 | 11 ± 1.4 | - | - | 93 ± 10.7 |
| Control | 2.1 ± 1.8 | < 0.005 | < 0.2 ± 1 | - | - | 2.5 ± 1.3 |

Mean values ± SD (n=3)

a, b, c) : see text for details

## EXAMPLE 12

### Identification of the Ste20p binding domain of Ste4p

Mutations within two regions of Ste4p which, when overexpressed, inhibited the signaling function of the wild-type protein were previously identified (Leberer et al., 1992, *supra*). The effect of two of these dominant-negative mutations within each region were examined for their effect on the association of Ste4p with either Ste20p or Ste5p. The K55E and D62N mutants of Ste4p (Leberer et al., 1992, *supra*) were defective in binding to GST-Ste20p, whereas binding to GST-Ste5p was normal (Fig. 3). The inability of these Ste4p mutants to bind Ste20p correlated with their sterile phenotype (Leberer et al., 1992, *supra*). However, the N157H/S175P and ΔF177 mutants which were also found to possess reduced signaling functions (Leberer et al., 1992, *supra*) were able to bind both Ste20p and Ste5p, although binding of Ste5p was reduced when compared with binding to wild-type Ste4p (Fig. 3), suggesting that this region may be involved in the interaction with an as yet unidentified component. The present invention therefore further provides means to identify this unidentified component and a further dissection of the structure-function relationship of Ste20p/PAK in signalling function.

Modeling of Ste4p by using the crystal structure of mammalian $G_{\beta 1}$ (Wall et al., 1995, Cell 83:1047-1058) as a template indicates that the residues predicted to interact with Ste20p are part of an amino-terminal α-helix in the region of $G_\beta$ that interacts with $G_\gamma$ (Wall et al., 1995, *supra*; Sondek et al., 1996, Nature, 379:369-374). The structure of yeast Ste4p ($G_\beta$) was modelled to the structure of mammalian $G_{\beta 1}$

(Wall et al., *supra*) using the homology module of Insight (Biosym, Inc.). Insertions specific for Ste4p were not considered.

However, consistent with the finding that the Ste4p$^{D62N}$ mutant interacted normally with Ste18p in the two-hybrid system (data not shown), the side chains of these residues are not predicted to be involved in the interaction with $G_\gamma$ but rather to be exposed on the cytoplasmic face of the $G_\beta$ structure (data not shown).

**Conclusion**

Together, these results indicate that transmission of the pheromone signal involves the regulated interaction between the mating-response G-protein β-subunit and a conserved sequence in the Ste20p protein kinase (Fig. 4). Pheromone-induced interaction with Ste4p may bring Ste20p in vicinity of Ste11p (Fig. 4) which interacts with Ste5p (Leberer et al., 1997, Curr. Op. Genet. & Dev., 7:59-66) and can serve as an *in vitro*-substrate for Ste20p (Wu et al., 1995, J. Biol. Chem., *supra*). Low concentrations of Ste20p/$G_\beta$ complexes present in the absence of pheromone may account for the basal signalling levels found in uninduced cells and may guarantee the rapid responsiveness of cells to pheromone (Chang et al., 1992, *supra*; Gartner et al., 1992, *supra*). In view of the high degree of conservation of Ste20p family protein kinases (Sells et al., 1997, *supra*), the results presented herein suggest that the interaction of these kinases with the β-subunit of heterotrimeric G-proteins (which are also highly conserved) may contribute to linking Ste20p homologs to G-protein-coupled receptors in other organisms including mammalian cells.

Although the present invention has been described hereinabove by way of preferred embodiments thereof, it can be modified, without departing from the spirit and nature of the subject invention as defined in the appended claims.

SEQUENCE LISTING

(1) GENERAL INFORMATION:

(i) APPLICANT: LEBERER, EKKEHARD
LEEUW, THOMAS
WHITEWAY, MALCOLM
THOMAS, DAVID Y.

(ii) TITLE OF INVENTION: THE G-PROTEIN BETA SUBUNIT INTERACTION DOMAIN OF STE20P/PAK FAMILY OF PROTEIN KINASES AND USES THEREOF IN BIOASSAYS

(iii) NUMBER OF SEQUENCES: 31

(iv) CORRESPONDENCE ADDRESS:
(A) ADDRESSEE: GOUDREAU GAGE DUBUC & MARTINEAU WALKER
(B) STREET: 3400 STOCK EXCHANGE TOWER, P.O. BOX 242, 800 VICTORIA SQUARE
(C) CITY: MONTREAL
(D) STATE: QUEBEC
(E) COUNTRY: CANADA
(F) ZIP: H4Z 1E9

(v) COMPUTER READABLE FORM:
(A) MEDIUM TYPE: Floppy disk
(B) COMPUTER: IBM PC compatible
(C) OPERATING SYSTEM: PC-DOS/MS-DOS
(D) SOFTWARE: PatentIn Release #1.0, Version #1.30

(vi) CURRENT APPLICATION DATA:
(A) APPLICATION NUMBER:
(B) FILING DATE:
(C) CLASSIFICATION:

(viii) ATTORNEY/AGENT INFORMATION:
(A) NAME: LECLERC, ALAIN M.
(B) REGISTRATION NUMBER: 37036
(C) REFERENCE/DOCKET NUMBER: 12219.9

(ix) TELECOMMUNICATION INFORMATION:
(A) TELEPHONE: 514-397-7675
(B) TELEFAX: 514-397-4382

(2) INFORMATION FOR SEQ ID NO:1:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 423 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: unknown
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

```
Met Ala Ala His Gln Met Asp Ser Ile Thr Tyr Ser Asn Asn Val Thr
1               5                   10                  15

Gln Gln Tyr Ile Gln Pro Gln Ser Leu Gln Asp Ile Ser Ala Val Glu
            20                  25                  30

Asp Glu Ile Gln Asn Lys Ile Glu Ala Ala Arg Gln Glu Ser Lys Gln
        35                  40                  45

Leu His Ala Gln Ile Asn Lys Ala Lys His Lys Ile Gln Asp Ala Ser
    50                  55                  60

Leu Phe Gln Met Ala Asn Lys Val Thr Ser Leu Thr Lys Asn Lys Ile
65                  70                  75                  80

Asn Leu Lys Pro Asn Ile Val Leu Lys Gly His Asn Asn Lys Ile Ser
                85                  90                  95

Asp Phe Arg Trp Ser Arg Asp Ser Lys Arg Ile Leu Ser Ala Ser Gln
            100                 105                 110

Asp Gly Phe Met Leu Ile Trp Asp Ser Ala Ser Gly Leu Lys Gln Asn
        115                 120                 125

Ala Ile Pro Leu Asp Ser Gln Trp Val Leu Ser Cys Ala Ile Ser Pro
    130                 135                 140

Ser Ser Thr Leu Val Ala Ser Ala Gly Leu Asn Asn Asn Cys Thr Ile
145                 150                 155                 160

Tyr Arg Val Ser Lys Glu Asn Arg Val Ala Gln Asn Val Ala Ser Ile
                165                 170                 175

Phe Lys Gly His Thr Cys Tyr Ile Ser Asp Ile Glu Phe Thr Asp Asn
            180                 185                 190

Ala His Ile Leu Thr Ala Ser Gly Asp Met Thr Cys Ala Leu Trp Asp
        195                 200                 205

Ile Pro Lys Ala Lys Arg Val Arg Glu Tyr Ser Asp His Leu Gly Asp
    210                 215                 220

Val Leu Ala Leu Ala Ile Pro Glu Glu Pro Asn Ser Glu Asn Ser Ser
225                 230                 235                 240

Asn Thr Phe Ala Ser Cys Gly Ser Asp Gly Tyr Thr Tyr Ile Trp Asp
                245                 250                 255

Ser Arg Ser Pro Ser Ala Val Gln Ser Phe Tyr Val Asn Asp Ser Asp
            260                 265                 270

Ile Asn Ala Leu Arg Phe Phe Lys Asp Gly Met Ser Ile Val Ala Gly
        275                 280                 285

Ser Asp Asn Gly Ala Ile Asn Met Tyr Asp Leu Arg Ser Asp Cys Ser
    290                 295                 300

Ile Ala Thr Phe Ser Leu Phe Arg Gly Tyr Glu Glu Arg Thr Pro Thr
```

```
305                 310                 315                 320

Pro Thr Tyr Met Ala Ala Asn Met Glu Tyr Asn Thr Ala Gln Ser Pro
                325                 330                 335

Gln Thr Leu Lys Ser Thr Ser Ser Ser Tyr Leu Asp Asn Gln Gly Val
            340                 345                 350

Val Ser Leu Asp Phe Ser Ala Ser Gly Arg Leu Met Tyr Ser Cys Tyr
        355                 360                 365

Thr Asp Ile Gly Cys Val Val Trp Asp Val Leu Lys Gly Glu Ile Val
    370                 375                 380

Gly Lys Leu Glu Gly His Gly Gly Arg Val Thr Gly Val Arg Ser Ser
385                 390                 395                 400

Pro Asp Gly Leu Ala Val Cys Thr Gly Ser Trp Asp Ser Thr Met Lys
                405                 410                 415

Ile Trp Ser Pro Gly Tyr Gln
            420
```

(2) INFORMATION FOR SEQ ID NO:2:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 340 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: unknown
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

```
Met Ser Glu Leu Asp Gln Leu Arg Gln Glu Ala Glu Gln Leu Lys Asn
1               5                   10                  15

Gln Ile Arg Asp Ala Arg Lys Ala Cys Ala Asp Ala Thr Leu Ser Gln
            20                  25                  30

Ile Thr Asn Asn Ile Asp Pro Val Gly Arg Ile Gln Met Arg Thr Arg
        35                  40                  45

Arg Thr Leu Arg Gly His Leu Ala Lys Ile Tyr Ala Met His Trp Gly
    50                  55                  60

Thr Asp Ser Arg Leu Leu Val Ser Ala Ser Gln Asp Gly Lys Leu Ile
65                  70                  75                  80

Ile Trp Asp Ser Tyr Thr Thr Asn Lys Val His Ala Ile Pro Leu Arg
                85                  90                  95

Ser Ser Trp Val Met Thr Cys Ala Tyr Ala Pro Ser Gly Asn Tyr Val
            100                 105                 110

Ala Cys Gly Gly Leu Asp Asn Ile Cys Ser Ile Tyr Asn Leu Lys Thr
        115                 120                 125
```

Arg Glu Gly Asn Val Arg Val Ser Arg Glu Leu Ala Gly His Thr Gly
130 135 140

Tyr Leu Ser Cys Cys Arg Phe Leu Asp Asp Asn Gln Ile Val Thr Ser
145 150 155 160

Ser Gly Asp Thr Thr Cys Ala Leu Trp Asp Ile Glu Thr Gly Gln Gln
165 170 175

Thr Thr Thr Phe Thr Gly His Thr Gly Asp Val Met Ser Leu Ser Leu
180 185 190

Ala Pro Asp Thr Arg Leu Phe Val Ser Gly Ala Cys Asp Ala Ser Ala
195 200 205

Lys Leu Trp Asp Val Arg Glu Gly Met Cys Arg Gln Thr Phe Thr Gly
210 215 220

His Glu Ser Asp Ile Asn Ala Ile Cys Phe Phe Pro Asn Gly Asn Ala
225 230 235 240

Phe Ala Thr Gly Ser Asp Asp Ala Thr Cys Arg Leu Phe Asp Leu Arg
245 250 255

Ala Asp Gln Glu Leu Met Thr Tyr Ser His Asp Asn Ile Ile Cys Gly
260 265 270

Ile Thr Ser Val Ser Phe Ser Lys Ser Gly Arg Leu Leu Leu Ala Gly
275 280 285

Tyr Asp Asp Phe Asn Cys Asn Val Trp Asp Ala Leu Lys Ala Asp Arg
290 295 300

Ala Gly Val Leu Ala Gly His Asp Asn Arg Val Ser Cys Leu Gly Val
305 310 315 320

Thr Asp Asp Gly Met Ala Val Ala Thr Gly Ser Trp Asp Ser Phe Leu
325 330 335

Lys Ile Trp Asn
340

(2) INFORMATION FOR SEQ ID NO:3:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 340 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: unknown
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

Met Ser Glu Leu Glu Gln Leu Arg Gln Glu Ala Glu Gln Leu Arg Asn
1 5 10 15

| | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Gln | Ile | Arg | Asp | Ala | Arg | Lys | Ala | Cys | Gly | Asp | Ser | Thr | Leu | Thr | Gln |
| | | | 20 | | | | | 25 | | | | | 30 | | |
| Ile | Thr | Ala | Gly | Leu | Asp | Pro | Val | Gly | Arg | Ile | Gln | Met | Arg | Thr | Arg |
| | | 35 | | | | | 40 | | | | | 45 | | | |
| Arg | Thr | Leu | Arg | Gly | His | Leu | Ala | Lys | Ile | Tyr | Ala | Met | His | Trp | Gly |
| | 50 | | | | | 55 | | | | | 60 | | | | |
| Thr | Asp | Ser | Arg | Leu | Leu | Val | Ser | Ala | Ser | Gln | Asp | Gly | Lys | Leu | Ile |
| 65 | | | | | 70 | | | | | 75 | | | | | 80 |
| Ile | Trp | Asp | Ser | Tyr | Thr | Thr | Asn | Lys | Val | His | Ala | Ile | Pro | Leu | Arg |
| | | | | 85 | | | | | 90 | | | | | 95 | |
| Ser | Ser | Trp | Val | Met | Thr | Cys | Ala | Tyr | Ala | Pro | Ser | Gly | Asn | Phe | Val |
| | | | 100 | | | | | 105 | | | | | 110 | | |
| Ala | Cys | Gly | Gly | Leu | Asp | Asn | Ile | Cys | Ser | Ile | Tyr | Ser | Leu | Lys | Thr |
| | | 115 | | | | | 120 | | | | | 125 | | | |
| Arg | Glu | Gly | Asn | Val | Arg | Val | Ser | Arg | Glu | Leu | Pro | Gly | His | Thr | Gly |
| | 130 | | | | | 135 | | | | | 140 | | | | |
| Tyr | Leu | Ser | Cys | Cys | Arg | Phe | Leu | Asp | Asp | Asn | Gln | Ile | Ile | Thr | Ser |
| 145 | | | | | 150 | | | | | 155 | | | | | 160 |
| Ser | Gly | Asp | Thr | Thr | Cys | Ala | Leu | Trp | Asp | Ile | Glu | Thr | Gly | Gln | Gln |
| | | | | 165 | | | | | 170 | | | | | 175 | |
| Thr | Val | Gly | Phe | Ala | Gly | His | Ser | Gly | Asp | Val | Met | Ser | Leu | Ser | Leu |
| | | | 180 | | | | | 185 | | | | | 190 | | |
| Ala | Pro | Asp | Gly | Arg | Thr | Phe | Val | Ser | Gly | Ala | Cys | Asp | Ala | Ser | Ile |
| | | 195 | | | | | 200 | | | | | 205 | | | |
| Lys | Leu | Trp | Asp | Val | Arg | Asp | Ser | Met | Cys | Arg | Gln | Thr | Phe | Ile | Gly |
| | 210 | | | | | 215 | | | | | 220 | | | | |
| His | Glu | Ser | Asp | Ile | Asn | Ala | Val | Ala | Phe | Phe | Pro | Asn | Gly | Tyr | Ala |
| 225 | | | | | 230 | | | | | 235 | | | | | 240 |
| Phe | Thr | Thr | Gly | Ser | Asp | Asp | Ala | Thr | Cys | Arg | Leu | Phe | Asp | Leu | Arg |
| | | | | 245 | | | | | 250 | | | | | 255 | |
| Ala | Asp | Gln | Glu | Leu | Leu | Met | Tyr | Ser | His | Asp | Asn | Ile | Ile | Cys | Gly |
| | | | 260 | | | | | 265 | | | | | 270 | | |
| Ile | Thr | Ser | Val | Ala | Phe | Ser | Arg | Ser | Gly | Arg | Leu | Leu | Leu | Ala | Gly |
| | | 275 | | | | | 280 | | | | | 285 | | | |
| Tyr | Asp | Asp | Phe | Asn | Cys | Asn | Ile | Trp | Asp | Ala | Met | Lys | Gly | Asp | Arg |
| | 290 | | | | | 295 | | | | | 300 | | | | |
| Ala | Gly | Val | Leu | Ala | Gly | His | Asp | Asn | Arg | Val | Ser | Cys | Leu | Gly | Val |
| 305 | | | | | 310 | | | | | 315 | | | | | 320 |

Thr Asp Asp Gly Met Ala Val Ala Thr Gly Ser Trp Asp Ser Phe Leu
325 330 335

Lys Ile Trp Asn
340

(2) INFORMATION FOR SEQ ID NO:4:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 340 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: unknown
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

Met Gly Glu Met Glu Gln Leu Arg Gln Glu Ala Glu Gln Leu Lys Lys
1 5 10 15

Gln Ile Ala Asp Ala Arg Lys Ala Cys Ala Asp Val Thr Leu Ala Glu
20 25 30

Leu Val Ser Gly Leu Glu Val Val Gly Arg Val Gln Met Arg Thr Arg
35 40 45

Arg Thr Leu Arg Gly His Leu Ala Lys Ile Tyr Ala Met His Trp Ala
50 55 60

Thr Asp Ser Lys Leu Leu Val Ser Ala Ser Gln Asp Gly Lys Leu Ile
65 70 75 80

Val Trp Asp Ser Tyr Thr Thr Asn Lys Val His Ala Ile Pro Leu Arg
85 90 95

Ser Ser Trp Val Met Thr Cys Ala Tyr Ala Pro Ser Gly Asn Phe Val
100 105 110

Ala Cys Gly Gly Leu Asp Asn Met Cys Ser Ile Tyr Asn Leu Lys Ser
115 120 125

Arg Glu Gly Asn Val Lys Val Ser Arg Glu Leu Ser Ala His Thr Gly
130 135 140

Tyr Leu Ser Cys Cys Arg Phe Leu Asp Asp Asn Asn Ile Val Thr Ser
145 150 155 160

Ser Gly Asp Thr Thr Cys Ala Leu Trp Asp Ile Glu Thr Gly Gln Gln
165 170 175

Lys Thr Val Phe Val Gly His Thr Gly Asp Cys Met Ser Leu Ala Val
180 185 190

Ser Pro Asp Phe Asn Leu Phe Ile Ser Gly Ala Cys Asp Ala Ser Ala
195 200 205

Lys Leu Trp Asp Val Arg Glu Gly Thr Cys Arg Gln Thr Phe Thr Gly
    210                 215                 220

His Glu Ser Asp Ile Asn Ala Ile Cys Phe Phe Pro Asn Gly Glu Ala
225                 230                 235                 240

Ile Cys Thr Gly Ser Asp Asp Ala Ser Cys Arg Leu Phe Asp Leu Arg
                245                 250                 255

Ala Asp Gln Glu Leu Ile Cys Phe Ser His Glu Ser Ile Ile Cys Gly
            260                 265                 270

Ile Thr Ser Val Ala Phe Ser Leu Ser Gly Arg Leu Leu Phe Ala Gly
        275                 280                 285

Tyr Asp Asp Phe Asn Cys Asn Val Trp Asp Ser Met Lys Ser Glu Arg
    290                 295                 300

Val Gly Ile Leu Ser Gly His Asp Asn Arg Val Ser Cys Leu Gly Val
305                 310                 315                 320

Thr Ala Asp Gly Met Ala Val Ala Thr Gly Ser Trp Asp Ser Phe Leu
                325                 330                 335

Lys Ile Trp Asn
            340

(2) INFORMATION FOR SEQ ID NO:5:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 340 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: unknown
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

Met Ser Glu Leu Glu Gln Leu Arg Gln Glu Ala Glu Gln Leu Arg Asn
1               5                   10                  15

Gln Ile Gln Asp Ala Arg Lys Ala Cys Asn Asp Ala Thr Leu Val Gln
            20                  25                  30

Ile Thr Ser Asn Met Asp Ser Val Gly Arg Ile Gln Met Arg Thr Arg
        35                  40                  45

Arg Thr Leu Arg Gly His Leu Ala Lys Ile Tyr Ala Met His Trp Gly
    50                  55                  60

Tyr Asp Ser Arg Leu Leu Val Ser Ala Ser Gln Asp Gly Lys Leu Ile
65                  70                  75                  80

Ile Trp Asp Ser Tyr Thr Thr Asn Lys Met His Ala Ile Pro Leu Arg
                85                  90                  95

```
Ser Ser Trp Val Met Thr Cys Ala Tyr Ala Pro Ser Gly Asn Tyr Val
            100                 105                 110

Ala Cys Gly Gly Leu Asp Asn Ile Cys Ser Ile Tyr Asn Leu Lys Thr
        115                 120                 125

Arg Glu Gly Asp Val Arg Val Ser Arg Glu Leu Ala Gly His Thr Gly
    130                 135                 140

Tyr Leu Ser Cys Cys Arg Phe Leu Asp Asp Gly Gln Ile Ile Thr Ser
145                 150                 155                 160

Ser Gly Asp Thr Thr Cys Ala Leu Trp Asp Ile Glu Thr Gly Gln Gln
                165                 170                 175

Thr Thr Thr Phe Thr Gly His Ser Gly Asp Val Met Ser Leu Ser Leu
            180                 185                 190

Ser Pro Asp Leu Lys Thr Phe Val Ser Gly Ala Cys Asp Ala Ser Ser
        195                 200                 205

Lys Leu Trp Asp Ile Arg Asp Gly Met Cys Arg Gln Ser Phe Thr Gly
    210                 215                 220

His Ile Ser Asp Ile Asn Ala Val Ser Phe Phe Pro Ser Gly Tyr Ala
225                 230                 235                 240

Phe Ala Thr Gly Ser Asp Asp Ala Thr Cys Arg Leu Phe Asp Leu Arg
                245                 250                 255

Ala Asp Gln Glu Leu Leu Leu Tyr Ser His Asp Asn Ile Ile Cys Gly
            260                 265                 270

Ile Thr Ser Val Ala Phe Ser Lys Ser Gly Arg Leu Leu Leu Ala Gly
        275                 280                 285

Tyr Asp Asp Phe Asn Cys Ser Val Trp Asp Ala Leu Lys Gly Gly Arg
    290                 295                 300

Ser Gly Val Leu Ala Gly His Asp Asn Arg Val Ser Cys Leu Gly Val
305                 310                 315                 320

Thr Asp Asp Gly Met Ala Val Ala Thr Gly Ser Trp Asp Ser Phe Leu
                325                 330                 335

Arg Ile Trp Asn
            340
```

(2) INFORMATION FOR SEQ ID NO:6:

(i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 353 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: unknown
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

| | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Met | Ala | Thr | Asp | Gly | Leu | His | Glu | Asn | Glu | Thr | Leu | Ala | Ser | Leu | Lys |
| 1 | | | | 5 | | | | | 10 | | | | | 15 | |
| Ser | Glu | Ala | Glu | Ser | Leu | Lys | Gly | Lys | Leu | Glu | Glu | Glu | Arg | Ala | Lys |
| | | | 20 | | | | | 25 | | | | | 30 | | |
| Leu | His | Asp | Val | Glu | Leu | His | Gln | Val | Ala | Glu | Arg | Val | Glu | Ala | Leu |
| | | 35 | | | | | 40 | | | | | 45 | | | |
| Gly | Gln | Phe | Val | Met | Lys | Thr | Arg | Arg | Thr | Leu | Lys | Gly | His | Gly | Asn |
| | 50 | | | | | 55 | | | | | 60 | | | | |
| Lys | Val | Leu | Cys | Met | Asp | Trp | Cys | Lys | Asp | Lys | Arg | Arg | Ile | Val | Ser |
| 65 | | | | | 70 | | | | | 75 | | | | | 80 |
| Ser | Ser | Gln | Asp | Gly | Lys | Val | Ile | Val | Trp | Asp | Ser | Phe | Thr | Thr | Asn |
| | | | | 85 | | | | | 90 | | | | | 95 | |
| Lys | Glu | His | Ala | Val | Thr | Met | Pro | Cys | Thr | Trp | Val | Met | Ala | Cys | Ala |
| | | | 100 | | | | | 105 | | | | | 110 | | |
| Tyr | Ala | Pro | Ser | Gly | Cys | Ala | Ile | Ala | Cys | Gly | Gly | Leu | Asp | Asn | Lys |
| | | 115 | | | | | 120 | | | | | 125 | | | |
| Cys | Ser | Val | Tyr | Pro | Leu | Thr | Phe | Asp | Lys | Asn | Glu | Asn | Met | Ala | Ala |
| | 130 | | | | | 135 | | | | | 140 | | | | |
| Lys | Lys | Lys | Ser | Val | Ala | Met | His | Thr | Asn | Tyr | Leu | Ser | Ala | Cys | Ser |
| 145 | | | | | 150 | | | | | 155 | | | | | 160 |
| Phe | Thr | Asn | Ser | Asp | Met | Gln | Ile | Leu | Thr | Ala | Ser | Gly | Asp | Gly | Thr |
| | | | | 165 | | | | | 170 | | | | | 175 | |
| Cys | Ala | Leu | Trp | Asp | Val | Glu | Ser | Gly | Gln | Leu | Leu | Gln | Ser | Phe | His |
| | | | 180 | | | | | 185 | | | | | 190 | | |
| Gly | His | Gly | Ala | Asp | Val | Leu | Cys | Leu | Asp | Leu | Ala | Pro | Ser | Glu | Thr |
| | | 195 | | | | | 200 | | | | | 205 | | | |
| Gly | Asn | Thr | Phe | Val | Ser | Gly | Gly | Cys | Asp | Lys | Lys | Ala | Met | Val | Trp |
| | 210 | | | | | 215 | | | | | 220 | | | | |
| Asp | Met | Arg | Ser | Gly | Gln | Cys | Val | Gln | Ala | Phe | Glu | Thr | His | Glu | Ser |
| 225 | | | | | 230 | | | | | 235 | | | | | 240 |
| Asp | Val | Asn | Ser | Val | Arg | Tyr | Tyr | Pro | Ser | Gly | Asp | Ala | Phe | Ala | Ser |
| | | | | 245 | | | | | 250 | | | | | 255 | |
| Gly | Ser | Asp | Asp | Ala | Thr | Cys | Arg | Leu | Tyr | Asp | Leu | Arg | Ala | Asp | Arg |
| | | | 260 | | | | | 265 | | | | | 270 | | |
| Glu | Val | Ala | Ile | Tyr | Ser | Lys | Glu | Ser | Ile | Ile | Phe | Gly | Ala | Ser | Ser |
| | | 275 | | | | | 280 | | | | | 285 | | | |

Val Asp Phe Ser Leu Ser Gly Arg Leu Leu Phe Ala Gly Tyr Asn Asp
    290                     295                     300

Tyr Thr Ile Asn Val Trp Asp Val Leu Lys Gly Ser Arg Val Ser Ile
305                 310                     315                     320

Leu Phe Gly His Glu Asn Arg Val Ser Thr Leu Arg Val Ser Pro Asp
                325                     330                     335

Gly Thr Ala Phe Cys Ser Gly Ser Trp Asp His Thr Leu Arg Val Trp
            340                     345                     350

Ala

(2) INFORMATION FOR SEQ ID NO:7:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 25 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: unknown
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

Glu Ala Asn Ser Ser Leu Ala Pro Leu Val Lys Leu Ala Arg Leu Lys
1               5                   10                  15

Lys Val Ala Glu Asn Met Asp Ala Asp
            20                  25

(2) INFORMATION FOR SEQ ID NO:8:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 22 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: unknown
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

Asp Asp Val Ser Ser Leu Ser Pro Leu Val Lys Ile Ala Arg Leu Lys
1               5                   10                  15

Lys Met Ser Glu Ser Asp
            20

(2) INFORMATION FOR SEQ ID NO:9:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 18 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: unknown
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:

Val Pro Val Ser Ser Leu Ile Pro Leu Ile Lys Ser Ile His His Ser
1 5 10 15

Gly Lys

(2) INFORMATION FOR SEQ ID NO:10:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 21 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: unknown
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:

Lys Pro Leu Ser Ser Leu Thr Pro Leu Ile Ala Ala Ala Lys Glu Ala
1 5 10 15

Thr Lys Asn Asn His
20

(2) INFORMATION FOR SEQ ID NO:11:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 21 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: unknown
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:

Lys Pro Leu Ser Ser Leu Thr Pro Leu Ile Met Ala Ala Lys Glu Ala
1 5 10 15

Met Lys Ser Asn Arg
20

(2) INFORMATION FOR SEQ ID NO:12:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 21 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: unknown
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:12:

```
Lys Pro Leu Ser Ser Leu Thr Pro Leu Ile Met Ala Ala Lys Glu Ala
1               5                   10                  15

Met Lys Ser Asn Arg
            20
```

(2) INFORMATION FOR SEQ ID NO:13:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 21 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: unknown
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:13:

```
Lys Pro Leu Ser Ser Leu Thr Pro Leu Ile Ala Ala Ala Lys Glu Ala
1               5                   10                  15

Thr Lys Asn Asn His
            20
```

(2) INFORMATION FOR SEQ ID NO:14:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 22 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: unknown
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:14:

```
Lys Pro Leu Ser Ser Leu Thr Pro Leu Ile Leu Ala Ala Lys Glu Ala
1               5                   10                  15

Ile Lys Asn Ser Ser Arg
            20
```

(2) INFORMATION FOR SEQ ID NO:15:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 21 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: unknown
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:15:

Lys Pro Leu Ser Ser Leu Thr Pro Leu Ile Leu Ala Ala Lys Glu Ala
1               5                   10                  15

Met Lys Ser Asn Arg
            20

(2) INFORMATION FOR SEQ ID NO:16:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 21 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: unknown
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:16:

Lys Pro Leu Ser Ser Leu Thr Pro Leu Ile Met Ala Ala Lys Glu Ala
1               5                   10                  15

Met Lys Ser Asn Arg
            20

(2) INFORMATION FOR SEQ ID NO:17:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 22 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: unknown
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:17:

Lys Pro Leu Ser Ser Leu Thr Pro Leu Ile Ile Ala Ala Lys Glu Ala
1               5                   10                  15

Ile Lys Asn Ser Ser Arg
            20

(2) INFORMATION FOR SEQ ID NO:18:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: unknown
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:18:

```
Arg Pro Leu Ala Ser Leu Thr Pro Leu Ile Met Ala Ala Lys Glu Ala
1               5                   10                  15

Thr Lys Gly Asn
            20
```

(2) INFORMATION FOR SEQ ID NO:19:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 21 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: unknown
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:19:

```
Lys Pro Leu Ser Ser Leu Thr Pro Tyr Ile Ile Thr Gly Lys Gln Ile
1               5                   10                  15

Ala Lys Gly Gly His
            20
```

(2) INFORMATION FOR SEQ ID NO:20:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 23 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: unknown
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:20:

```
Lys Pro Leu Ala Ser Leu Tyr Tyr Leu Ile Val Ala Ala Lys Lys Ser
1               5                   10                  15

Ile Ala Glu Ala Ser Asn Ser
            20
```

(2) INFORMATION FOR SEQ ID NO:21:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: unknown
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:21:

```
Cys Asn Cys Asn Gly Leu Val Pro Ala Ile Met Glu Ala Lys Lys Ala
1               5                   10                  15
```

Lys Glu Ala His Ser Lys Phe Ser Ile His
20 25

(2) INFORMATION FOR SEQ ID NO:22:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 25 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: unknown
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:22:

Gly Pro Glu Ser Asp Leu Ile Pro Leu Val Glu Arg Thr Lys Asn Glu
1 5 10 15

Ala Gln Arg Asp Phe Ser Met Phe Phe
20 25

(2) INFORMATION FOR SEQ ID NO:23:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 14 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: unknown
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:23:

Cys Asp Pro Lys Asp Leu Thr Ser Leu Leu Glu Trp Lys Glu
1 5 10

(2) INFORMATION FOR SEQ ID NO:24:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 32 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: unknown
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:24:

Gly Lys Ile Glu Glu Leu Ala Pro Leu Leu Glu Trp Lys Lys Gln Gln
1 5 10 15

Gln Lys His Gln Gln His Lys Gln Glu Thr Ser Asp Thr Gly Phe Ala
20 25 30

(2) INFORMATION FOR SEQ ID NO:25:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 13 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: unknown
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:25:

```
Cys Ser Pro Glu Gln Leu Lys Val Ser Leu Lys Trp His
1               5                   10
```

(2) INFORMATION FOR SEQ ID NO:26:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: unknown
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:26:

```
Cys Pro Thr Glu Asp Leu Lys Ser Ile Ile Phe Ser Arg Lys Ala Asn
1               5                   10                  15

Thr His Ile Asn
            20
```

(2) INFORMATION FOR SEQ ID NO:27:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 11 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: unknown
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:27:

```
Ser Ser Leu Xaa Pro Leu Xaa Xaa Xaa Xaa Xaa
1               5                   10
```

(2) INFORMATION FOR SEQ ID NO:28:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 4136 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: unknown
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 276..3092

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:28:

```
GAATTCGAAA GTCTACCGCT TTTGGCAGCT GAAAAATTCA GAAAGTCACC TGGCCAGAGA       60

GGAAAAATAC GAAACCAAAA AGAGGCATCC GTAAATTCGC ATTAGCAACG CATGCTTACA      120

TAGATACTCA CATACTACAC ACACTTACAT ACTTTCTTAA AGACATACAT CCGTACGTAC      180

AATTAGAGCG AGGTAGCAAG CAACCCAAAC TTCTTCCCTT CACTGCCTCA CACCCCATCC      240

TAAATATCCC ACAAGATCCT CGACTAATAC AAGAA ATG AGC AAT GAT CCA TCT         293
                                      Met Ser Asn Asp Pro Ser
                                        1               5

GCT GTA TCG GAA CTA CCA GAC AAG GAC AGT CTT GAT AAC GGT ATC AGC        341
Ala Val Ser Glu Leu Pro Asp Lys Asp Ser Leu Asp Asn Gly Ile Ser
             10                  15                  20

AAT GAC AAT GAA AGG GCC ATG GGC GGC AAT GGC GAT GGC GGC GAT GGA        389
Asn Asp Asn Glu Arg Ala Met Gly Gly Asn Gly Asp Gly Gly Asp Gly
         25                  30                  35

TTA CGA TTA CCA AGG ACC ACT GGA ACT TTG AAC GTC AAT GCC TTA CAA        437
Leu Arg Leu Pro Arg Thr Thr Gly Thr Leu Asn Val Asn Ala Leu Gln
     40                  45                  50

AAA GGC ACT AAT GCT GCC CAT GAA GCT GGT GGA TAC AAA TCC ATG GAT        485
Lys Gly Thr Asn Ala Ala His Glu Ala Gly Gly Tyr Lys Ser Met Asp
 55                  60                  65                  70

CCT GCG AAG AAC GCG GAG ACA ACC AAT GAT GAT GAC AAT AAT GTC GTT        533
Pro Ala Lys Asn Ala Glu Thr Thr Asn Asp Asp Asp Asn Asn Val Val
                 75                  80                  85

TCA CTA GAT GAT CCT ATT CAA TTT ACC CGA GTA TCT TCC TCC TCT GTC        581
Ser Leu Asp Asp Pro Ile Gln Phe Thr Arg Val Ser Ser Ser Ser Val
             90                  95                 100

ATC AGT GGA ATG TCT TCA TCC ATG AGT CCT CAT TCT AAC ATC GAT GAA        629
Ile Ser Gly Met Ser Ser Ser Met Ser Pro His Ser Asn Ile Asp Glu
        105                 110                 115

ACC AAA TCT CTA GAA GCA GTC ACT CCA AAC ATA AAT ACC AGC AAT ATA        677
Thr Lys Ser Leu Glu Ala Val Thr Pro Asn Ile Asn Thr Ser Asn Ile
    120                 125                 130

ACC CCG GAT CAT TCT GCT GAC AAC ACA TTT TCT ACC ATA AAT GCG TCC        725
Thr Pro Asp His Ser Ala Asp Asn Thr Phe Ser Thr Ile Asn Ala Ser
135                 140                 145                 150

GAG TCA GAT CAC CAG TTT AAT GAC ACT TTA CTA TCA AAA CTG TCG TTA        773
Glu Ser Asp His Gln Phe Asn Asp Thr Leu Leu Ser Lys Leu Ser Leu
                155                 160                 165
```

```
ACA GAT TCT ACA GAA ACT ATA GAA AAT AAC GCG ACA GTG AAG CAC CAG      821
Thr Asp Ser Thr Glu Thr Ile Glu Asn Asn Ala Thr Val Lys His Gln
            170                 175                 180

CAG CCA GTT GCA TCT TCC ACA GTA AAC TCG AAT AAG AGC TCC ACT GAT      869
Gln Pro Val Ala Ser Ser Thr Val Asn Ser Asn Lys Ser Ser Thr Asp
        185                 190                 195

ATA CGA AGG GCT ACA CCA GTG TCC ACT CCC GTT ATC TCT AAA CCA TCG      917
Ile Arg Arg Ala Thr Pro Val Ser Thr Pro Val Ile Ser Lys Pro Ser
    200                 205                 210

ATG ACA ACC ACG CCA AGA CAG ATC AAT TCA GCT TCC CAT TCG CTT TCG      965
Met Thr Thr Thr Pro Arg Gln Ile Asn Ser Ala Ser His Ser Leu Ser
215                 220                 225                 230

AAC CCT AAG CAT AAG CAA CAT AAA CCA AAA GTT AAA CCG TCC AAG CCT     1013
Asn Pro Lys His Lys Gln His Lys Pro Lys Val Lys Pro Ser Lys Pro
                235                 240                 245

GAA GCA AAA AGT AAA CCG GTT TCT GTG AAA AAA AGC TTT CCT TCG AAA     1061
Glu Ala Lys Ser Lys Pro Val Ser Val Lys Lys Ser Phe Pro Ser Lys
            250                 255                 260

AAT CCT TTA AAA AAC TCC TCT CCA CCT AAA AAG CAA ACA GAA AAA TCG     1109
Asn Pro Leu Lys Asn Ser Ser Pro Pro Lys Lys Gln Thr Glu Lys Ser
        265                 270                 275

TAT TAT TCT TCC TCT TCG AAA AAA AGG AAA AGC GGT TCA AAT AGT GGT     1157
Tyr Tyr Ser Ser Ser Ser Lys Lys Arg Lys Ser Gly Ser Asn Ser Gly
    280                 285                 290

ACA CTA AGA ATG AAA GAT GTC TTT ACG TCC TTT GTA CAG AAT ATA AAG     1205
Thr Leu Arg Met Lys Asp Val Phe Thr Ser Phe Val Gln Asn Ile Lys
295                 300                 305                 310

AGA AAT TCT CAG GAT GAT AAA AGG GCC TCA TCG TCG TCC AAT AAT TCT     1253
Arg Asn Ser Gln Asp Asp Lys Arg Ala Ser Ser Ser Ser Asn Asn Ser
                315                 320                 325

TCC TCA TCT TCT ATA ACC ACC GCT TTG AGG ATA TCT ACG CCA TAC AAT     1301
Ser Ser Ser Ser Ile Thr Thr Ala Leu Arg Ile Ser Thr Pro Tyr Asn
            330                 335                 340

GCC AAG CAT ATC CAC CAT GTG GGC GTG GAC TCC AAG ACT GGT GAG TAC     1349
Ala Lys His Ile His His Val Gly Val Asp Ser Lys Thr Gly Glu Tyr
        345                 350                 355

ACA GGT TTG CCG GAG GAA TGG GAA AAA TTG TTG ACT TCT AGT GGT ATT     1397
Thr Gly Leu Pro Glu Glu Trp Glu Lys Leu Leu Thr Ser Ser Gly Ile
    360                 365                 370

TCC AAA AGA GAA CAA CAG CAA AAC ATG CAA GCA GTC ATG GAT ATT GTC     1445
Ser Lys Arg Glu Gln Gln Gln Asn Met Gln Ala Val Met Asp Ile Val
375                 380                 385                 390
```

```
AAA TTC TAT CAG GAT GTC ACG GAA ACA AAC GGT GAA GAT AAA ATG TTC       1493
Lys Phe Tyr Gln Asp Val Thr Glu Thr Asn Gly Glu Asp Lys Met Phe
                395                 400                 405

AAG ACT TTC AAC ACA ACC ACA GGA TTG CCG GGA AGT CCT CAA GTT TCA       1541
Lys Thr Phe Asn Thr Thr Thr Gly Leu Pro Gly Ser Pro Gln Val Ser
            410                 415                 420

ACA CCG CCT GCA AAC TCA TTC AAT AAA TTT CCT CCG TCG ACA AGT GAT       1589
Thr Pro Pro Ala Asn Ser Phe Asn Lys Phe Pro Pro Ser Thr Ser Asp
        425                 430                 435

TCG CAC AAT TAC GGT TCC AGA ACA GGT ACA CCA ATG TCC AAT CAC GTC       1637
Ser His Asn Tyr Gly Ser Arg Thr Gly Thr Pro Met Ser Asn His Val
    440                 445                 450

ATG TCT CCA ACC TTA AAT ACA GAT TCT AGT TCA GCA AAC GGG AAA TTC       1685
Met Ser Pro Thr Leu Asn Thr Asp Ser Ser Ser Ala Asn Gly Lys Phe
455                 460                 465                 470

ATA CCA AGT AGA CCG GCT CCT AAG CCC CCA TCT TCT GCG TCC GCT TCA       1733
Ile Pro Ser Arg Pro Ala Pro Lys Pro Pro Ser Ser Ala Ser Ala Ser
                475                 480                 485

GCT CCA ATT ATA AAA TCA CCC GTC ATG AAT TCT GCC GCC AAT GTT TCG       1781
Ala Pro Ile Ile Lys Ser Pro Val Met Asn Ser Ala Ala Asn Val Ser
            490                 495                 500

CCC TTG AAG CAG ACT CAT GCA CCT ACA ACT CCG AAC AGG ACC AGC CCA       1829
Pro Leu Lys Gln Thr His Ala Pro Thr Thr Pro Asn Arg Thr Ser Pro
        505                 510                 515

AAC AGG TCC TCA ATA TCA AGA AAT GCC ACT TTA AAA AAA GAG GAG CAG       1877
Asn Arg Ser Ser Ile Ser Arg Asn Ala Thr Leu Lys Lys Glu Glu Gln
    520                 525                 530

CCA CTA CCA CCA ATA CCT CCA ACC AAA TCC AAA ACG TCT CCA ATC ATC       1925
Pro Leu Pro Pro Ile Pro Pro Thr Lys Ser Lys Thr Ser Pro Ile Ile
535                 540                 545                 550

TCC ACA GCT CAC ACA CCA CAG CAA GTT GCT CAA TCG CCA AAA GCG CCG       1973
Ser Thr Ala His Thr Pro Gln Gln Val Ala Gln Ser Pro Lys Ala Pro
                555                 560                 565

GCG CAA GAG ACG GTA ACG ACA CCT ACT TCG AAG CCA GCT CAA GCA AGA       2021
Ala Gln Glu Thr Val Thr Thr Pro Thr Ser Lys Pro Ala Gln Ala Arg
            570                 575                 580

AGC TTG TCT AAA GAA TTA AAT GAG AAA AAG AGA GAG GAA AGG GAA AGA       2069
Ser Leu Ser Lys Glu Leu Asn Glu Lys Lys Arg Glu Glu Arg Glu Arg
        585                 590                 595

CGT AAA AAA CAA CTA TAT GCC AAA TTG AAC GAA ATT TGC TCA GAC GGT       2117
Arg Lys Lys Gln Leu Tyr Ala Lys Leu Asn Glu Ile Cys Ser Asp Gly
    600                 605                 610
```

```
GAC CCA AGT ACA AAA TAT GCC AAT TTA GTA AAA ATT GGT CAA GGT GCA    2165
Asp Pro Ser Thr Lys Tyr Ala Asn Leu Val Lys Ile Gly Gln Gly Ala
615                 620                 625                 630

TCA GGT GGT GTT TAT ACT GCT TAT GAA ATA GGT ACG AAT GTC TCA GTG    2213
Ser Gly Gly Val Tyr Thr Ala Tyr Glu Ile Gly Thr Asn Val Ser Val
                635                 640                 645

GCC ATT AAG CAA ATG AAT CTC GAA AAG CAA CCA AAA AAG GAG CTA ATC    2261
Ala Ile Lys Gln Met Asn Leu Glu Lys Gln Pro Lys Lys Glu Leu Ile
            650                 655                 660

ATC AAT GAG ATT CTG GTC ATG AAG GGT AGC AAA CAC CCT AAT ATA GTT    2309
Ile Asn Glu Ile Leu Val Met Lys Gly Ser Lys His Pro Asn Ile Val
        665                 670                 675

AAT TTC ATT GAT TCT TAC GTT TTA AAA GGC GAC CTT TGG GTC ATT ATG    2357
Asn Phe Ile Asp Ser Tyr Val Leu Lys Gly Asp Leu Trp Val Ile Met
    680                 685                 690

GAA TAC ATG GAA GGT GGC TCC TTA ACT GAT GTG GTC ACC CAT TGT ATT    2405
Glu Tyr Met Glu Gly Gly Ser Leu Thr Asp Val Val Thr His Cys Ile
695                 700                 705                 710

TTG ACA GAA GGT CAA ATT GGT GCC GTT TGT AGA GAA ACT TTG AGT GGG    2453
Leu Thr Glu Gly Gln Ile Gly Ala Val Cys Arg Glu Thr Leu Ser Gly
                715                 720                 725

TTG GAA TTT TTA CAT TCT AAA GGT GTT CTT CAC AGA GAT ATC AAA TCC    2501
Leu Glu Phe Leu His Ser Lys Gly Val Leu His Arg Asp Ile Lys Ser
            730                 735                 740

GAT AAC ATC CTA TTG TCC ATG GAA GGG GAT ATT AAG TTA ACG GAT TTC    2549
Asp Asn Ile Leu Leu Ser Met Glu Gly Asp Ile Lys Leu Thr Asp Phe
        745                 750                 755

GGT TTT TGC GCT CAA ATC AAT GAA TTG AAC TTG AAA AGA ACT ACT ATG    2597
Gly Phe Cys Ala Gln Ile Asn Glu Leu Asn Leu Lys Arg Thr Thr Met
    760                 765                 770

GTG GGA ACG CCT TAT TGG ATG GCG CCT GAA GTG GTT TCT AGG AAA GAA    2645
Val Gly Thr Pro Tyr Trp Met Ala Pro Glu Val Val Ser Arg Lys Glu
775                 780                 785                 790

TAT GGC CCA AAA GTA GAT ATC TGG TCG TTG GGT ATC ATG ATC ATT GAA    2693
Tyr Gly Pro Lys Val Asp Ile Trp Ser Leu Gly Ile Met Ile Ile Glu
                795                 800                 805

ATG ATC GAG GGG GAG CCT CCA TAT TTA AAT GAA ACC CCG CTA AGA GCA    2741
Met Ile Glu Gly Glu Pro Pro Tyr Leu Asn Glu Thr Pro Leu Arg Ala
            810                 815                 820

CTG TAT TTA ATT GCT ACA AAT GGT ACA CCC AAG TTA AAG GAA CCC GAG    2789
Leu Tyr Leu Ile Ala Thr Asn Gly Thr Pro Lys Leu Lys Glu Pro Glu
        825                 830                 835
```

```
AAT CTA TCG TCA AGC TTG AAA AAA TTC CTT GAT TGG TGT TTA TGT GTG      2837
Asn Leu Ser Ser Ser Leu Lys Lys Phe Leu Asp Trp Cys Leu Cys Val
    840                 845                 850

GAG CCC GAA GAT AGA GCA AGC GCT ACG GAA TTG CTT CAT GAT GAA TAT      2885
Glu Pro Glu Asp Arg Ala Ser Ala Thr Glu Leu Leu His Asp Glu Tyr
855                 860                 865                 870

ATC ACG GAG ATA GCT GAA GCC AAT TCC TCA TTG GCC CCG CTA GTC AAG      2933
Ile Thr Glu Ile Ala Glu Ala Asn Ser Ser Leu Ala Pro Leu Val Lys
                875                 880                 885

TTA GCA AGA TTG AAG AAA GTA GCT GAG AAC ATG GAT GCT GAT GAA GAT      2981
Leu Ala Arg Leu Lys Lys Val Ala Glu Asn Met Asp Ala Asp Glu Asp
            890                 895                 900

AAT GAC GAC GAT AAC GAC AAC GAG CAT ATT AAT AAG ACA AAC AAT TGT      3029
Asn Asp Asp Asp Asn Asp Asn Glu His Ile Asn Lys Thr Asn Asn Cys
        905                 910                 915

GAC GAC AAT AAC GAT AGC AAA GAA ACC GTA AAT TTG GAC GTA ACT GAA      3077
Asp Asp Asn Asn Asp Ser Lys Glu Thr Val Asn Leu Asp Val Thr Glu
    920                 925                 930

GAT GAT AAA CAA AAG TAAACGTAGC AAGCAGGGTA CACCTTATTA TCGACAAAGT      3132
Asp Asp Lys Gln Lys
935

ATATACACAG TTGTGACTGG CATAAAAATT CTTTTCATAT ATCTTATCGT GTATATTTGG    3192

ACATTTTATA ACACATCCCA CTCTAATTCA CAACTTCATT AACGAAATTT AAATAAATCA    3252

CGACAACAGT TTTGCTTAAA ACTGAGGAAT ATTGAAACCA ACTCAAATTC TTCCTAATTT    3312

CAGGCGTATA AAAATAACAA ATTCTCATCG ATTGTCGGGT ACCATTACAC GAACATCTGT    3372

CTGCGTTCTA TGTAACGAAG GAGAGGTATT ATCCAATTTT GGAAATATCC GTAATATTGT    3432

CCTTAGTGCA CGAACTATAT TATCCCGCAA ATTCAGGGAA AAGAAAAGAA GTAGAAAAAA    3492

AAAAATACCA TGGGAGTCAG TTCTTGTTCA GCTGAGAGAA TTACGCTTGT TTCTTATTTC    3552

CCACATATAC GAGAAATTCC TACCGATATA ACATCCTCTC TCGTCTTCTA GAATTTTCCA    3612

GTTGAGTGAA GTTTTTTATT TTCATAAACT AACAAGATTA TTTCATGGAA CAGTGACGGA    3672

AAGGATTTTC TAAAGGCATT GTTAGAAAAA ATGGTTGACG ACTCAAACTA TCTTACACCA    3732

CATGAAACTG CATTAGCGGT GGTGGCCACT GCAATGAAGA AAGCAAGACT GCAACTAGAT    3792

ACATTGCTAA TAAATTCCAT ACTTGGTGGC GTTCTGTTTA GTAGTGGTTC GTTCTATTGG    3852

TAGCGGTATA TTCCGAAGAT CCTGACATAG TCGCACGAAA CCCGGGTATT GTGAATCTTA    3912

TTACTGGTGT TAATTTCGCC ATGGGACTAT TCTATGTAGT AATGATGGGT GCTGACCTCT    3972

TCAACTCTAA TATCCTATTT TTCTCCGTTG GAGTTCTGAG AAAAGCAGTA ACTATCTATG    4032
```

```
ATTTGATGAT TTCGTGGGTT GTCAGTTGGT TAGGTAATAT TGCTGGCTCA CTTTTTGTTT    4092

CATATCTTTT TGGTCATCTT TCTGGTATTA GTTCTCAGAA GCTT                     4136
```

(2) INFORMATION FOR SEQ ID NO:29:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 939 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:29:

```
Met Ser Asn Asp Pro Ser Ala Val Ser Glu Leu Pro Asp Lys Asp Ser
 1               5                  10                  15

Leu Asp Asn Gly Ile Ser Asn Asp Asn Glu Arg Ala Met Gly Gly Asn
            20                  25                  30

Gly Asp Gly Gly Asp Gly Leu Arg Leu Pro Arg Thr Thr Gly Thr Leu
        35                  40                  45

Asn Val Asn Ala Leu Gln Lys Gly Thr Asn Ala Ala His Glu Ala Gly
    50                  55                  60

Gly Tyr Lys Ser Met Asp Pro Ala Lys Asn Ala Glu Thr Thr Asn Asp
65                  70                  75                  80

Asp Asp Asn Asn Val Val Ser Leu Asp Asp Pro Ile Gln Phe Thr Arg
                85                  90                  95

Val Ser Ser Ser Ser Val Ile Ser Gly Met Ser Ser Ser Met Ser Pro
            100                 105                 110

His Ser Asn Ile Asp Glu Thr Lys Ser Leu Glu Ala Val Thr Pro Asn
        115                 120                 125

Ile Asn Thr Ser Asn Ile Thr Pro Asp His Ser Ala Asp Asn Thr Phe
    130                 135                 140

Ser Thr Ile Asn Ala Ser Glu Ser Asp His Gln Phe Asn Asp Thr Leu
145                 150                 155                 160

Leu Ser Lys Leu Ser Leu Thr Asp Ser Thr Glu Thr Ile Glu Asn Asn
                165                 170                 175

Ala Thr Val Lys His Gln Gln Pro Val Ala Ser Ser Thr Val Asn Ser
            180                 185                 190

Asn Lys Ser Ser Thr Asp Ile Arg Arg Ala Thr Pro Val Ser Thr Pro
        195                 200                 205

Val Ile Ser Lys Pro Ser Met Thr Thr Thr Pro Arg Gln Ile Asn Ser
    210                 215                 220
```

Ala Ser His Ser Leu Ser Asn Pro Lys His Lys Gln His Lys Pro Lys
225 230 235 240

Val Lys Pro Ser Lys Pro Glu Ala Lys Ser Lys Pro Val Ser Val Lys
245 250 255

Lys Ser Phe Pro Ser Lys Asn Pro Leu Lys Asn Ser Ser Pro Pro Lys
260 265 270

Lys Gln Thr Glu Lys Ser Tyr Tyr Ser Ser Ser Ser Lys Lys Arg Lys
275 280 285

Ser Gly Ser Asn Ser Gly Thr Leu Arg Met Lys Asp Val Phe Thr Ser
290 295 300

Phe Val Gln Asn Ile Lys Arg Asn Ser Gln Asp Asp Lys Arg Ala Ser
305 310 315 320

Ser Ser Ser Asn Asn Ser Ser Ser Ser Ser Ile Thr Thr Ala Leu Arg
325 330 335

Ile Ser Thr Pro Tyr Asn Ala Lys His Ile His His Val Gly Val Asp
340 345 350

Ser Lys Thr Gly Glu Tyr Thr Gly Leu Pro Glu Glu Trp Glu Lys Leu
355 360 365

Leu Thr Ser Ser Gly Ile Ser Lys Arg Glu Gln Gln Gln Asn Met Gln
370 375 380

Ala Val Met Asp Ile Val Lys Phe Tyr Gln Asp Val Thr Glu Thr Asn
385 390 395 400

Gly Glu Asp Lys Met Phe Lys Thr Phe Asn Thr Thr Thr Gly Leu Pro
405 410 415

Gly Ser Pro Gln Val Ser Thr Pro Pro Ala Asn Ser Phe Asn Lys Phe
420 425 430

Pro Pro Ser Thr Ser Asp Ser His Asn Tyr Gly Ser Arg Thr Gly Thr
435 440 445

Pro Met Ser Asn His Val Met Ser Pro Thr Leu Asn Thr Asp Ser Ser
450 455 460

Ser Ala Asn Gly Lys Phe Ile Pro Ser Arg Pro Ala Pro Lys Pro Pro
465 470 475 480

Ser Ser Ala Ser Ala Ser Ala Pro Ile Ile Lys Ser Pro Val Met Asn
485 490 495

Ser Ala Ala Asn Val Ser Pro Leu Lys Gln Thr His Ala Pro Thr Thr
500 505 510

Pro Asn Arg Thr Ser Pro Asn Arg Ser Ser Ile Ser Arg Asn Ala Thr
515 520 525

| | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Leu | Lys | Lys | Glu | Glu | Gln | Pro | Leu | Pro | Pro | Ile | Pro | Pro | Thr | Lys | Ser |
| | 530 | | | | | 535 | | | | | 540 | | | | |
| Lys | Thr | Ser | Pro | Ile | Ile | Ser | Thr | Ala | His | Thr | Pro | Gln | Gln | Val | Ala |
| 545 | | | | | 550 | | | | | 555 | | | | | 560 |
| Gln | Ser | Pro | Lys | Ala | Pro | Ala | Gln | Glu | Thr | Val | Thr | Thr | Pro | Thr | Ser |
| | | | | 565 | | | | | 570 | | | | | 575 | |
| Lys | Pro | Ala | Gln | Ala | Arg | Ser | Leu | Ser | Lys | Glu | Leu | Asn | Glu | Lys | Lys |
| | | | 580 | | | | | 585 | | | | | 590 | | |
| Arg | Glu | Glu | Arg | Glu | Arg | Arg | Lys | Lys | Gln | Leu | Tyr | Ala | Lys | Leu | Asn |
| | | 595 | | | | | 600 | | | | | 605 | | | |
| Glu | Ile | Cys | Ser | Asp | Gly | Asp | Pro | Ser | Thr | Lys | Tyr | Ala | Asn | Leu | Val |
| | 610 | | | | | 615 | | | | | 620 | | | | |
| Lys | Ile | Gly | Gln | Gly | Ala | Ser | Gly | Gly | Val | Tyr | Thr | Ala | Tyr | Glu | Ile |
| 625 | | | | | 630 | | | | | 635 | | | | | 640 |
| Gly | Thr | Asn | Val | Ser | Val | Ala | Ile | Lys | Gln | Met | Asn | Leu | Glu | Lys | Gln |
| | | | | 645 | | | | | 650 | | | | | 655 | |
| Pro | Lys | Lys | Glu | Leu | Ile | Ile | Asn | Glu | Ile | Leu | Val | Met | Lys | Gly | Ser |
| | | | 660 | | | | | 665 | | | | | 670 | | |
| Lys | His | Pro | Asn | Ile | Val | Asn | Phe | Ile | Asp | Ser | Tyr | Val | Leu | Lys | Gly |
| | | 675 | | | | | 680 | | | | | 685 | | | |
| Asp | Leu | Trp | Val | Ile | Met | Glu | Tyr | Met | Glu | Gly | Gly | Ser | Leu | Thr | Asp |
| | 690 | | | | | 695 | | | | | 700 | | | | |
| Val | Val | Thr | His | Cys | Ile | Leu | Thr | Glu | Gly | Gln | Ile | Gly | Ala | Val | Cys |
| 705 | | | | | 710 | | | | | 715 | | | | | 720 |
| Arg | Glu | Thr | Leu | Ser | Gly | Leu | Glu | Phe | Leu | His | Ser | Lys | Gly | Val | Leu |
| | | | | 725 | | | | | 730 | | | | | 735 | |
| His | Arg | Asp | Ile | Lys | Ser | Asp | Asn | Ile | Leu | Leu | Ser | Met | Glu | Gly | Asp |
| | | | 740 | | | | | 745 | | | | | 750 | | |
| Ile | Lys | Leu | Thr | Asp | Phe | Gly | Phe | Cys | Ala | Gln | Ile | Asn | Glu | Leu | Asn |
| | | 755 | | | | | 760 | | | | | 765 | | | |
| Leu | Lys | Arg | Thr | Thr | Met | Val | Gly | Thr | Pro | Tyr | Trp | Met | Ala | Pro | Glu |
| | 770 | | | | | 775 | | | | | 780 | | | | |
| Val | Val | Ser | Arg | Lys | Glu | Tyr | Gly | Pro | Lys | Val | Asp | Ile | Trp | Ser | Leu |
| 785 | | | | | 790 | | | | | 795 | | | | | 800 |
| Gly | Ile | Met | Ile | Ile | Glu | Met | Ile | Glu | Gly | Glu | Pro | Pro | Tyr | Leu | Asn |
| | | | | 805 | | | | | 810 | | | | | 815 | |
| Glu | Thr | Pro | Leu | Arg | Ala | Leu | Tyr | Leu | Ile | Ala | Thr | Asn | Gly | Thr | Pro |
| | | | 820 | | | | | 825 | | | | | 830 | | |

```
Lys Leu Lys Glu Pro Glu Asn Leu Ser Ser Ser Leu Lys Lys Phe Leu
        835                 840                 845

Asp Trp Cys Leu Cys Val Glu Pro Glu Asp Arg Ala Ser Ala Thr Glu
    850                 855                 860

Leu Leu His Asp Glu Tyr Ile Thr Glu Ile Ala Glu Ala Asn Ser Ser
865                 870                 875                 880

Leu Ala Pro Leu Val Lys Leu Ala Arg Leu Lys Lys Val Ala Glu Asn
                885                 890                 895

Met Asp Ala Asp Glu Asp Asn Asp Asp Asp Asn Asp Asn Glu His Ile
            900                 905                 910

Asn Lys Thr Asn Asn Cys Asp Asp Asn Asn Asp Ser Lys Glu Thr Val
        915                 920                 925

Asn Leu Asp Val Thr Glu Asp Asp Lys Gln Lys
    930                 935
```

(2) INFORMATION FOR SEQ ID NO:30:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2325 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: unknown
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 496..1764

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:30:

```
GAAAAATGTT TCAGGAAGAG ATACTGCGTA AAAAAAAGAC ACATGTGTTA CGCAGGAAAA    60

AGTTTGTGAG GCTTTTTGCC TTAACAGATT GACTTGTAGC CCTGTTAGGT TTACCCAACA   120

TTTGTTTTTC TGTGTGTCGA AAATTTTTTC AGAGTGTTTT CAACTGACAC TTGCCTGTTT   180

CATATTAGTT GTAACTTAAA CTTTCAAACA TAAAACTTTT TTGGAAGTCC ATCCTTCACA   240

TGACTTGAAT CCCTTCAATA TCGAAACAGT TATCCTCAAA ATCTCTTATC ACTTTTCTAA   300

TTGTTTTCTT CCCCTTTTTT GTAGTAACTC GCTGTAAAGC ACATTTTATT CATAATCTCC   360

TTTGTGCCAG AACTCAAGGT CAATAGGCCA GAATTATTGG AAGGAAAGAG GGAAGAAAAT   420

ACGATATTGC TAGTTCATTA AGTCAAGGAA GAAAATACTC AAAAAACTGT ACAGCTCAAT   480

CAGGTACACA TTACG ATG GCA GCA CAT CAG ATG GAC TCG ATA ACG TAT TCT    531
                 Met Ala Ala His Gln Met Asp Ser Ile Thr Tyr Ser
                  1               5                  10
```

```
AAT AAT GTC ACC CAA CAG TAT ATA CAA CCA CAA AGT CTA CAG GAT ATC      579
Asn Asn Val Thr Gln Gln Tyr Ile Gln Pro Gln Ser Leu Gln Asp Ile
         15                  20                  25

TCT GCA GTG GAG GAA GAA ATT CAA AAT AAA ATA GAG GCC GCC AGA CAA      627
Ser Ala Val Glu Glu Glu Ile Gln Asn Lys Ile Glu Ala Ala Arg Gln
     30                  35                  40

GAG AGT AAA CAG CTT CAT GCT CAA ATA AAT AAA GCA AAA CAC AAG ATA      675
Glu Ser Lys Gln Leu His Ala Gln Ile Asn Lys Ala Lys His Lys Ile
 45                  50                  55                  60

CAA GAT GCA AGC TTA TTC CAG ATG GCC AAC AAA GTT ACT TCG TTG ACC      723
Gln Asp Ala Ser Leu Phe Gln Met Ala Asn Lys Val Thr Ser Leu Thr
                 65                  70                  75

AAA AAT AAG ATC AAC TTA AAG CCA AAT ATC GTG TTG AAA GGC CAT AAT      771
Lys Asn Lys Ile Asn Leu Lys Pro Asn Ile Val Leu Lys Gly His Asn
             80                  85                  90

AAT AAA ATC TCA GAT TTT CGG TGG AGT CGA GAT TCA AAA CGT ATT TTG      819
Asn Lys Ile Ser Asp Phe Arg Trp Ser Arg Asp Ser Lys Arg Ile Leu
         95                 100                 105

AGT GCA AGT CAA GAT GGC TTT ATG CTT ATA TGG GAC AGT GCT TCA GGT      867
Ser Ala Ser Gln Asp Gly Phe Met Leu Ile Trp Asp Ser Ala Ser Gly
    110                 115                 120

TTA AAA CAG AAC GCT ATT CCA TTA GAT TCT CAA TGG GTT CTT TCC TGC      915
Leu Lys Gln Asn Ala Ile Pro Leu Asp Ser Gln Trp Val Leu Ser Cys
125                 130                 135                 140

GCT ATT TCG CCA TCG AGT ACT TTG GTA GCA AGC GCA GGA TTA AAC AAT      963
Ala Ile Ser Pro Ser Ser Thr Leu Val Ala Ser Ala Gly Leu Asn Asn
                145                 150                 155

AAC TGT ACC ATT TAT AGA GTT TCG AAA GAA AAC AGA GTA GCG CAA AAC     1011
Asn Cys Thr Ile Tyr Arg Val Ser Lys Glu Asn Arg Val Ala Gln Asn
            160                 165                 170

GTT GCG TCA ATT TTC AAA GGA CAT ACT TGC TAT ATT TCT GAC ATT GAA     1059
Val Ala Ser Ile Phe Lys Gly His Thr Cys Tyr Ile Ser Asp Ile Glu
        175                 180                 185

TTT ACA GAT AAC GCA CAT ATA TTG ACA GCA AGT GGG GAT ATG ACA TGT     1107
Phe Thr Asp Asn Ala His Ile Leu Thr Ala Ser Gly Asp Met Thr Cys
    190                 195                 200

GCC TTG TGG GAT ATA CCG AAA GCA AAG AGG GTG AGA GAA TAT TCT GAC     1155
Ala Leu Trp Asp Ile Pro Lys Ala Lys Arg Val Arg Glu Tyr Ser Asp
205                 210                 215                 220

CAT TTA GGT GAT GTT TTG GCA TTA GCT ATT CCT GAA GAG CCA AAC TTA     1203
His Leu Gly Asp Val Leu Ala Leu Ala Ile Pro Glu Glu Pro Asn Leu
                225                 230                 235
```

GAA AAT TCT TCG AAC ACA TTC GCT AGC TGT GGA TCA GAC GGG TAT ACT 1251
Glu Asn Ser Ser Asn Thr Phe Ala Ser Cys Gly Ser Asp Gly Tyr Thr
240 245 250

TAC ATA TGG GAT AGC AGA TCT CCG TCC GCT GTA CAA AGC TTT TAC GTT 1299
Tyr Ile Trp Asp Ser Arg Ser Pro Ser Ala Val Gln Ser Phe Tyr Val
255 260 265

AAC GAT AGT GAT ATT AAT GCA CTT CGT TTT TTC AAA GAC GGG ATG TCG 1347
Asn Asp Ser Asp Ile Asn Ala Leu Arg Phe Phe Lys Asp Gly Met Ser
270 275 280

ATT GTT GCA GGA AGT GAC AAT GGT GCG ATA AAT ATG TAT GAT TTA AGG 1395
Ile Val Ala Gly Ser Asp Asn Gly Ala Ile Asn Met Tyr Asp Leu Arg
285 290 295 300

TCG GAC TGT TCT ATT GCT ACT TTT TCT CTT TTT CGA GGT TAT GAA GAA 1443
Ser Asp Cys Ser Ile Ala Thr Phe Ser Leu Phe Arg Gly Tyr Glu Glu
305 310 315

CGT ACC CCT ACC CCT ACT TAT ATG GCA GCT AAC ATG GAG TAC AAT ACC 1491
Arg Thr Pro Thr Pro Thr Tyr Met Ala Ala Asn Met Glu Tyr Asn Thr
320 325 330

GCG CAA TCG CCA CAA ACT TTA AAA TCA ACA AGC TCA AGC TAT CTA GAC 1539
Ala Gln Ser Pro Gln Thr Leu Lys Ser Thr Ser Ser Ser Tyr Leu Asp
335 340 345

AAC CAA GGC GTT GTT TCT TTA GAT TTT AGT GCA TCT GGA AGA TTG ATG 1587
Asn Gln Gly Val Val Ser Leu Asp Phe Ser Ala Ser Gly Arg Leu Met
350 355 360

TAC TCA TGC TAT ACA GAC ATT GGT TGT GTT GTG TGG GAT GTA TTA AAA 1635
Tyr Ser Cys Tyr Thr Asp Ile Gly Cys Val Val Trp Asp Val Leu Lys
365 370 375 380

GGA GAG ATT GTT GGA AAA TTA GAA GGT CAT GGT GGC AGA GTC ACT GGT 1683
Gly Glu Ile Val Gly Lys Leu Glu Gly His Gly Gly Arg Val Thr Gly
385 390 395

GTG CGC TCG AGT CCA GAT GGG TTA GCT GTA TGT ACA GGT TCA TGG GAC 1731
Val Arg Ser Ser Pro Asp Gly Leu Ala Val Cys Thr Gly Ser Trp Asp
400 405 410

TCA ACC ATG AAA ATA TGG TCT CCA GGT TAT CAA TAGCTTCGAA TTGGAAATAC 1784
Ser Thr Met Lys Ile Trp Ser Pro Gly Tyr Gln
415 420

TGTGAGCAGT AATTATCAAT GGATGCTATT ATATAAATAT ACATACCTAC ACCCATCCCA 1844

TATTTACATA GAATAACAAC AGTAACATTA GTTCTGTGGA AGCGCAAAAA CGTCCTTTAA 1904

TAAAGTAAGT CAAAACATTC AACAATGAAA ATTCAAAGCA TTGTCATTTG CTTCCTTTTT 1964

CTCTTTGGGA TAAACGAAAC AAAAACGAAC AAAATGTCAT GCACTCAAAA ATTCTTTTCA 2024

ATCGTTTTGG AAACAGTATT ATTCACTGAC TTATTTGACC AACTTGCTAG AATCATCTAT 2084

GTTTTCAGGC ATTGTTTAAT TTCATGATGG CTGTCCCTAC TTTAGCTTGT TATGAGCCTT 2144

CACTGGCTCG TCCTTATGTA TTGCGTCTGA CCCAAAATTT GTCCTTTCTT GTTTAGTGGA 2204

ATTTTTGTTC GGTAATTTCA AAAATGCTGA ATTTTGATTA ACAAATCATC TGGTAGTTGT 2264

GTTATAAACA TAAAAAACTG CTCCCTTCTG GGATGATTTT CAATTGCTCT CTGTACTGCA 2324

G 2325

(2) INFORMATION FOR SEQ ID NO:31:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 423 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:31:

Met Ala Ala His Gln Met Asp Ser Ile Thr Tyr Ser Asn Asn Val Thr
1 5 10 15

Gln Gln Tyr Ile Gln Pro Gln Ser Leu Gln Asp Ile Ser Ala Val Glu
20 25 30

Glu Glu Ile Gln Asn Lys Ile Glu Ala Ala Arg Gln Glu Ser Lys Gln
35 40 45

Leu His Ala Gln Ile Asn Lys Ala Lys His Lys Ile Gln Asp Ala Ser
50 55 60

Leu Phe Gln Met Ala Asn Lys Val Thr Ser Leu Thr Lys Asn Lys Ile
65 70 75 80

Asn Leu Lys Pro Asn Ile Val Leu Lys Gly His Asn Asn Lys Ile Ser
85 90 95

Asp Phe Arg Trp Ser Arg Asp Ser Lys Arg Ile Leu Ser Ala Ser Gln
100 105 110

Asp Gly Phe Met Leu Ile Trp Asp Ser Ala Ser Gly Leu Lys Gln Asn
115 120 125

Ala Ile Pro Leu Asp Ser Gln Trp Val Leu Ser Cys Ala Ile Ser Pro
130 135 140

Ser Ser Thr Leu Val Ala Ser Ala Gly Leu Asn Asn Asn Cys Thr Ile
145 150 155 160

Tyr Arg Val Ser Lys Glu Asn Arg Val Ala Gln Asn Val Ala Ser Ile
165 170 175

Phe Lys Gly His Thr Cys Tyr Ile Ser Asp Ile Glu Phe Thr Asp Asn
180 185 190

```
Ala His Ile Leu Thr Ala Ser Gly Asp Met Thr Cys Ala Leu Trp Asp
        195                 200                 205

Ile Pro Lys Ala Lys Arg Val Arg Glu Tyr Ser Asp His Leu Gly Asp
    210                 215                 220

Val Leu Ala Leu Ala Ile Pro Glu Glu Pro Asn Leu Glu Asn Ser Ser
225                 230                 235                 240

Asn Thr Phe Ala Ser Cys Gly Ser Asp Gly Tyr Thr Tyr Ile Trp Asp
                245                 250                 255

Ser Arg Ser Pro Ser Ala Val Gln Ser Phe Tyr Val Asn Asp Ser Asp
            260                 265                 270

Ile Asn Ala Leu Arg Phe Phe Lys Asp Gly Met Ser Ile Val Ala Gly
        275                 280                 285

Ser Asp Asn Gly Ala Ile Asn Met Tyr Asp Leu Arg Ser Asp Cys Ser
    290                 295                 300

Ile Ala Thr Phe Ser Leu Phe Arg Gly Tyr Glu Glu Arg Thr Pro Thr
305                 310                 315                 320

Pro Thr Tyr Met Ala Ala Asn Met Glu Tyr Asn Thr Ala Gln Ser Pro
                325                 330                 335

Gln Thr Leu Lys Ser Thr Ser Ser Ser Tyr Leu Asp Asn Gln Gly Val
            340                 345                 350

Val Ser Leu Asp Phe Ser Ala Ser Gly Arg Leu Met Tyr Ser Cys Tyr
        355                 360                 365

Thr Asp Ile Gly Cys Val Val Trp Asp Val Leu Lys Gly Glu Ile Val
    370                 375                 380

Gly Lys Leu Glu Gly His Gly Gly Arg Val Thr Gly Val Arg Ser Ser
385                 390                 395                 400

Pro Asp Gly Leu Ala Val Cys Thr Gly Ser Trp Asp Ser Thr Met Lys
                405                 410                 415

Ile Trp Ser Pro Gly Tyr Gln
            420
```

WHAT IS CLAIMED IS:

1. An isolated Ste4p/$G_\beta$-binding polypeptide or fragment thereof wherein said isolated Ste4p/$G_\beta$-binding polypeptide is a Ste20p/PAK polypeptide which directly binds to a Ste4p/$G_\beta$ polypeptide or fragment thereof.

2. The isolated Ste4p/$G_\beta$-binding polypeptide of claim 1, comprising an amino acid sequence having at least 95 % identity to the amino acid sequence selected from the group consisting of:

a) a full length amino acid sequence of SEQ. ID. NO.:29;

b) an amino acid sequence having amino acids 495 to 939 of SEQ. ID. NO.:29;

c) an amino acid sequence having amino acids 495 to 888 of SEQ. ID. NO.:29;

d) an amino acid sequence having amino acids 819 to 939 of SEQ. ID. NO.:29;

e) an amino acid sequence having amino acids 819 to 892 of SEQ. ID. NO.:29;

f) an amino acid sequence having amino acids 876 to 939 of SEQ. ID. NO.:29; and

g) an amino acid sequence having amino acids 876 to 892 of SEQ. ID. NO.:29.

3. The isolated Ste4p/$G_\beta$-binding polypeptide of claim 1, comprising an amino acid sequence having at least 95 % identity to the amino acid sequence selected from the group consisting of:

a) an amino acid sequence of SEQ. ID. NO.:29;
b) an amino acid sequence of SEQ. ID. NO.:1;
c) an amino acid sequence of SEQ. ID. NO.:2;
d) an amino acid sequence of SEQ. ID. NO.:3;
e) an amino acid sequence of SEQ. ID. NO.:4;
f) an amino acid sequence of SEQ. ID. NO.:5;
g) an amino acid sequence of SEQ. ID. NO.:6;
h) an amino acid sequence of SEQ. ID. NO.:7;
i) an amino acid sequence of SEQ. ID. NO.:8;
j) an amino acid sequence of SEQ. ID. NO.:9;
k) an amino acid sequence of SEQ. ID. NO.:10;
l) an amino acid sequence of SEQ. ID. NO.:11;
m) an amino acid sequence of SEQ. ID. NO.:12; and
n) an amino acid sequence of SEQ. ID. NO.:13.

4. The isolated Ste4p/$G_\beta$-binding polypeptide of claim 1, joined to a heterologous polypeptide, thereby forming an isolated chimeric polypeptide which directly binds to a Ste4p/$G_\beta$ polypeptide or fragment thereof.

5. The isolated Ste4p/$G_\beta$-binding polypeptide of claim 4, comprising amino acid sequence SSL$\phi$PL$I_v$X$\phi\phi\beta$, wherein $\phi$ is selected from A, I, L, M, S, T, and $\beta$ is selected from H, K, and R.

6. The isolated Ste4p/$G_\beta$-binding polypeptide of claim 5, wherein said heterologous polypeptide is Glutathione-S-transferase.

7. An isolated nucleic acid molecule comprising a nucleic acid sequence which encodes a Ste4p/$G_\beta$ binding domain of Ste20p/PAK.

8. The isolated nucleic acid of claim 7, comprising a nucleic acid sequence at least 90 % identical to a sequence selected from the group consisting of:

a) a nucleotide sequence encoding a full length amino acid sequence of SEQ. ID. NO.:29;

b) a nucleotide sequence encoding an amino acid sequence having amino acid 495 to 939 of SEQ. ID. NO.:29;

c) a nucleotide sequence encoding an amino acid sequence having amino acid 495 to 888 of SEQ. ID. NO.:29;

d) a nucleotide sequence encoding an amino acid sequence having amino acid 819 to 939 of SEQ. ID. NO.:29;

e) a nucleotide sequence encoding an amino acid sequence having amino acid 819 to 892 of SEQ. ID. NO.:29;

f) a nucleotide sequence encoding an amino acid sequence having amino acid 876 to 939 of SEQ. ID. NO.:29;

g) a nucleotide sequence encoding an amino acid sequence having amino acid 876 to 892 of SEQ. ID. NO.:29; and

h) a nucleotide sequence which hybridizes to a) - g) under high stringency conditions.

9. The isolated nucleic acid of claim 7, comprising a nucleic acid sequence at least 90 % identical to a sequence selected from the group consisting of:

a) a nucleotide sequence encoding an amino acid sequence of SEQ. ID. NO.:29;

b) a nucleotide sequence encoding an amino acid sequence of SEQ. ID. NO.:1;

c) a nucleotide sequence encoding an amino acid sequence of SEQ. ID. NO.:2;

d) a nucleotide sequence encoding an amino acid sequence of SEQ. ID. NO.:3;

e) a nucleotide sequence encoding an amino acid sequence of SEQ. ID. NO.:4;

f) a nucleotide sequence encoding an amino acid sequence of SEQ. ID. NO.:5;

g) a nucleotide sequence encoding an amino acid sequence of SEQ. ID. NO.:6;

h) a nucleotide sequence encoding an amino acid sequence of SEQ. ID. NO.:7;

i) a nucleotide sequence encoding an amino acid sequence of SEQ. ID. NO.:8;

j) a nucleotide sequence encoding an amino acid sequence of SEQ. ID. NO.:9;

k) a nucleotide sequence encoding an amino acid sequence of SEQ. ID. NO.:10;

l) a nucleotide sequence encoding an amino acid sequence of SEQ. ID. NO.:11;

m) a nucleotide sequence encoding an amino acid sequence of SEQ. ID. NO.:12; and

n) a nucleotide sequence encoding an amino acid sequence of SEQ. ID. NO.:13;

o) a nucleotide sequence encoding an amino acid sequence SSL$\phi$PLI$_v$X$\phi\phi\beta$, wherein $\phi$ is selected from A, I, L, M, S, T, and $\beta$ is selected from H, K, and R;

p) a nucleotide sequence which hybridizes to a) - e) under high stringency conditions.

10. An isolated nucleic acid molecule encoding a Ste4p/$G_\beta$ binding domain containing a fusion protein, said Ste4p/G $\beta$ interaction domain being fused to a heterologous polypeptide sequence, wherein said Ste4p/$G_\beta$ binding domain is encoded by the nucleic acid molecule of claim 7.

11. The isolated nucleic acid molecule of claim 10 wherein said Ste4p/$G_\beta$ binding domain comprises amino acid sequence SSL$\phi$PLI$_v$X$\phi\phi\beta$, wherein $\phi$ is selected from A, I, L, M, S, T, and $\beta$ is selected from H, K, and R.

12. The isolated nucleic acid molecule of claim 11, wherein said heterologous polypeptide sequence encodes Glutathione-S-transferase.

13. An isolated Ste20p/PAK-binding polypeptide or fragment thereof, wherein said isolated Ste20p/PAK polypeptide is a Ste4p/$G_\beta$ polypeptide which directly binds to a Ste20p/PAK polypeptide or fragment thereof.

14. The isolated Ste20p/PAK-binding polypeptide of claim 13, comprising an amino acid sequence having at least 95 % identity to the amino acid sequence selected from the group consisting of:

a) a full length amino acid sequence of SEQ. ID. NO.:21;

b) an amino acid sequence having amino acids 1 to 150 of SEQ. ID. NO.:21;

c) an amino acid sequence having amino acids 1 to 100 of SEQ. ID. NO.:21;

d) an amino acid sequence having amino acids 1 to 80 of SEQ. ID. NO.:21;

e) an amino acid sequence of SEQ. ID. NO.:22;

f) an amino acid sequence of SEQ. ID. NO.:23;

g) an amino acid sequence of SEQ. ID. NO.:24;

h) an amino acid sequence of SEQ. ID. NO.:25; and

i) an amino acid sequence of SEQ. ID. NO.:26.

15. The isolated Ste20p/PAK-binding polypeptide or fragment of claim 13, joined to a heterologous polypeptide, thereby forming a chimeric polypeptide which directly binds to a Ste20p/PAK polypeptide or fragment thereof.

16. The isolated Ste20p/PAK-binding polypeptide or fragment of claim 15, wherein said heterologous polypeptide is the influenza hemagglutinin (HA) - epitope.

17. An isolated nucleic acid molecule comprising a nucleic acid sequence which encodes a Ste20p/PAK binding domain of Ste4p/$G_\beta$ according to claim 13.

18. A vector comprising the nucleic acid molecule of claim 7.

19. A vector comprising the nucleic acid molecule of claim 17.

20. A host cell harboring the nucleic acid molecule of claim 18.

21. A host cell harboring the nucleic acid molecule of claim 19.

22. The host cell of claim 20 further harboring a vector comprising a nucleic acid molecule which comprises a nucleic acid sequence encoding a Ste20p/PAK interaction domain of Ste4p/$G_\beta$.

23. The host cell of claim 22, wherein at least one assayable metabolic function is dependent on the interaction of said nucleic acid sequences encoding the Ste4p/$G_\beta$ interaction domain of Ste20p/PAK and the Ste20p/PAK interaction domain of Ste4p/$G_\beta$.

24. A method of assaying compounds having the ability to modulate the interaction between Ste20p/PAK and Ste4p/$G_\beta$ comprising the steps of:

a) incubating the host cell of claim 23 with a test compound;

b) assaying said at least one metabolic function dependent on said interaction between Ste20p/PAK and Ste4p/$G_\beta$; and

c) identifying said compound as a modulator of said interaction.

25. A method of assaying compounds having the ability to modulate the interaction between Ste20p/PAK and Ste4p/$G_\beta$ comprising the steps of:

a) incubating the isolated Ste4p/$G_\beta$ polypeptide of claim 5 with an isolated Ste20p/PAK-binding polypeptide or fragment thereof, wherein said isolated Ste20p/PAK-binding polypeptide is a

Ste4p/$G_\beta$ polypeptide which directly binds to a Ste20p/PAK polypeptide or fragment thereof, in the presence of a test compound;

b) assaying *in vitro* the binding between said isolated Ste4p/$G_\beta$ polypeptide and said isolated Ste20p/PAK polypeptide; and

c) identifying said compound as a modulator of said interaction.

26. A composition of matter comprising:

a) an isolated Ste4p/$G_\beta$-binding polypeptide or fragment thereof wherein said isolated Ste4p/$G_\beta$-binding polypeptide is a Ste20p/PAK polypeptide which directly binds to a Ste4p/$G_\beta$ polypeptide or fragment thereof; and

b) an isolated Ste20p/PAK-binding polypeptide or fragment thereof, wherein said isolated Ste20p/PAK polypeptide is a Ste4p/$G_\beta$ polypeptide which directly binds to a Ste20p/PAK polypeptide or fragment thereof.

27. A composition of matter comprising an isolated nucleic acid molecule comprising a nucleic acid sequence which encodes a Ste4p/$G_\beta$-binding domain of Ste20p/PAK and an isolated nucleic acid molecule comprising a nucleic acid sequence which encodes a Ste20p/PAK-binding domain of Ste4p/$G_\beta$.

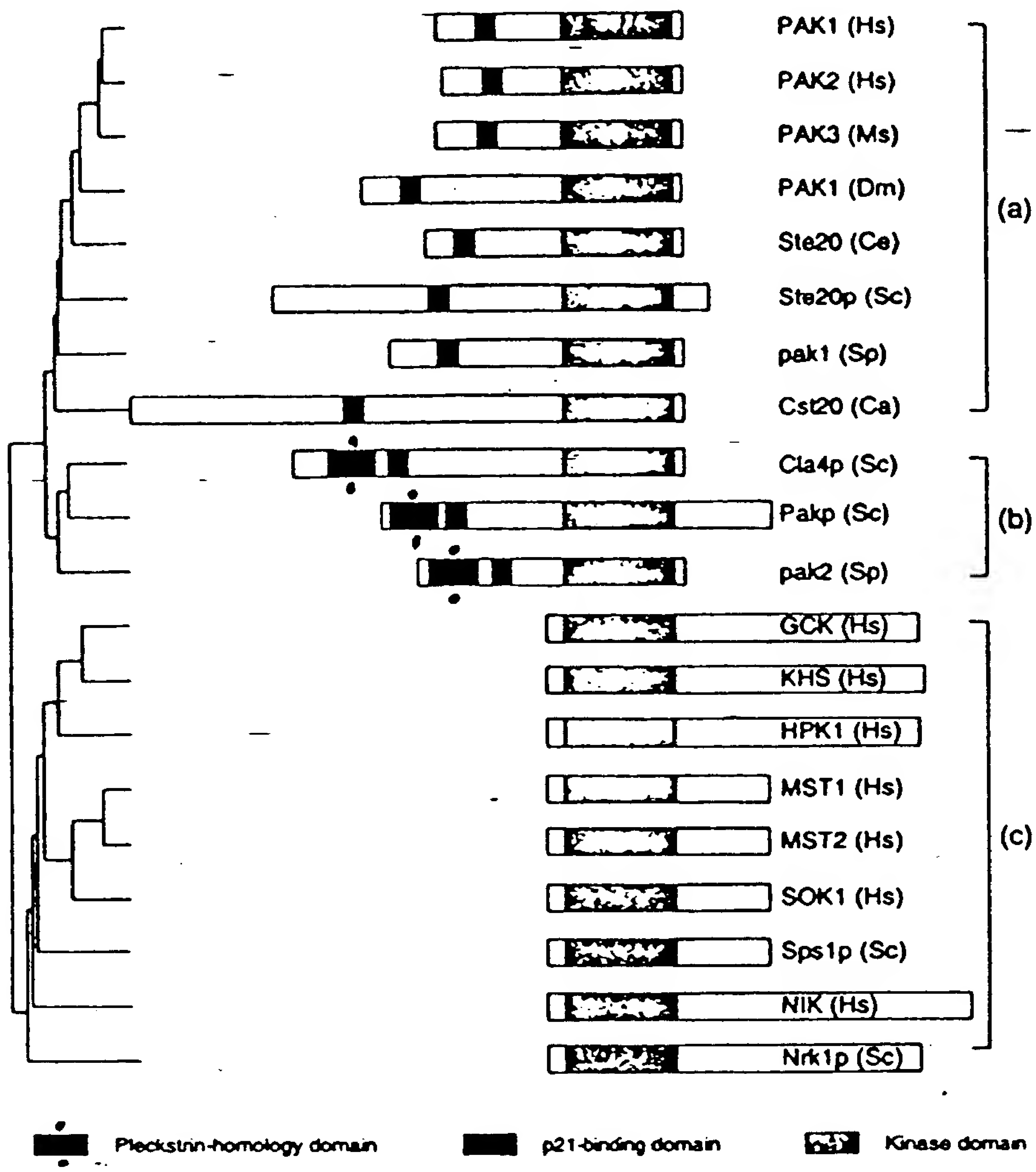


FIGURE A ( Prior Art)

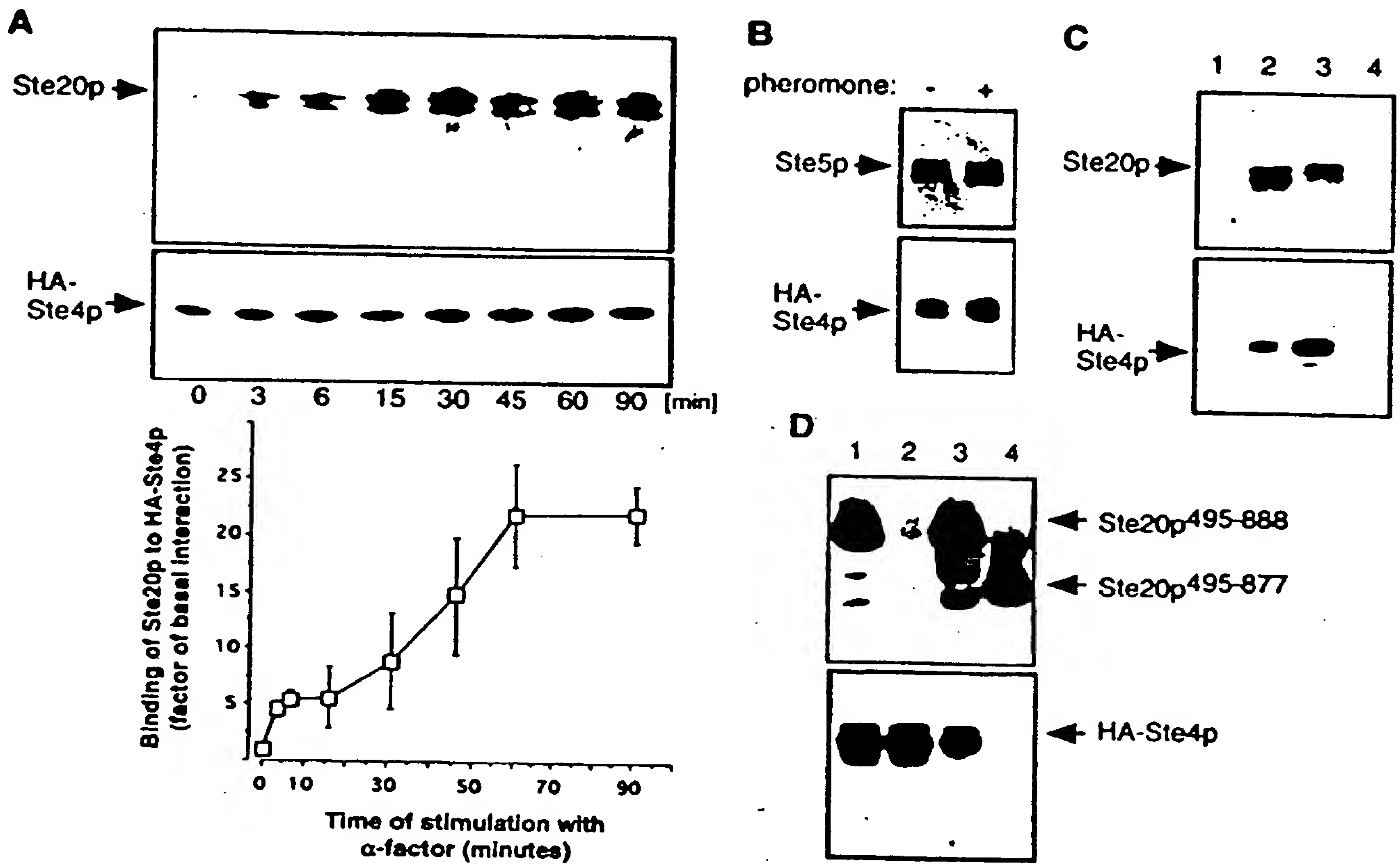
A
Ste20p
HA-Ste4p
0 3 6 15 30 45 60 90 [min]
Binding of Ste20p to HA-Ste4p (factor of basal interaction)
25
20
15
10
5
0 10 30 50 70 90
Time of stimulation with α-factor (minutes)
B
pheromone: - +
Ste5p
HA-Ste4p
C
1 2 3 4
Ste20p
HA-Ste4p
D
1 2 3 4
Ste20p495-888
Ste20p495-877
HA-Ste4p

FIGURE 1

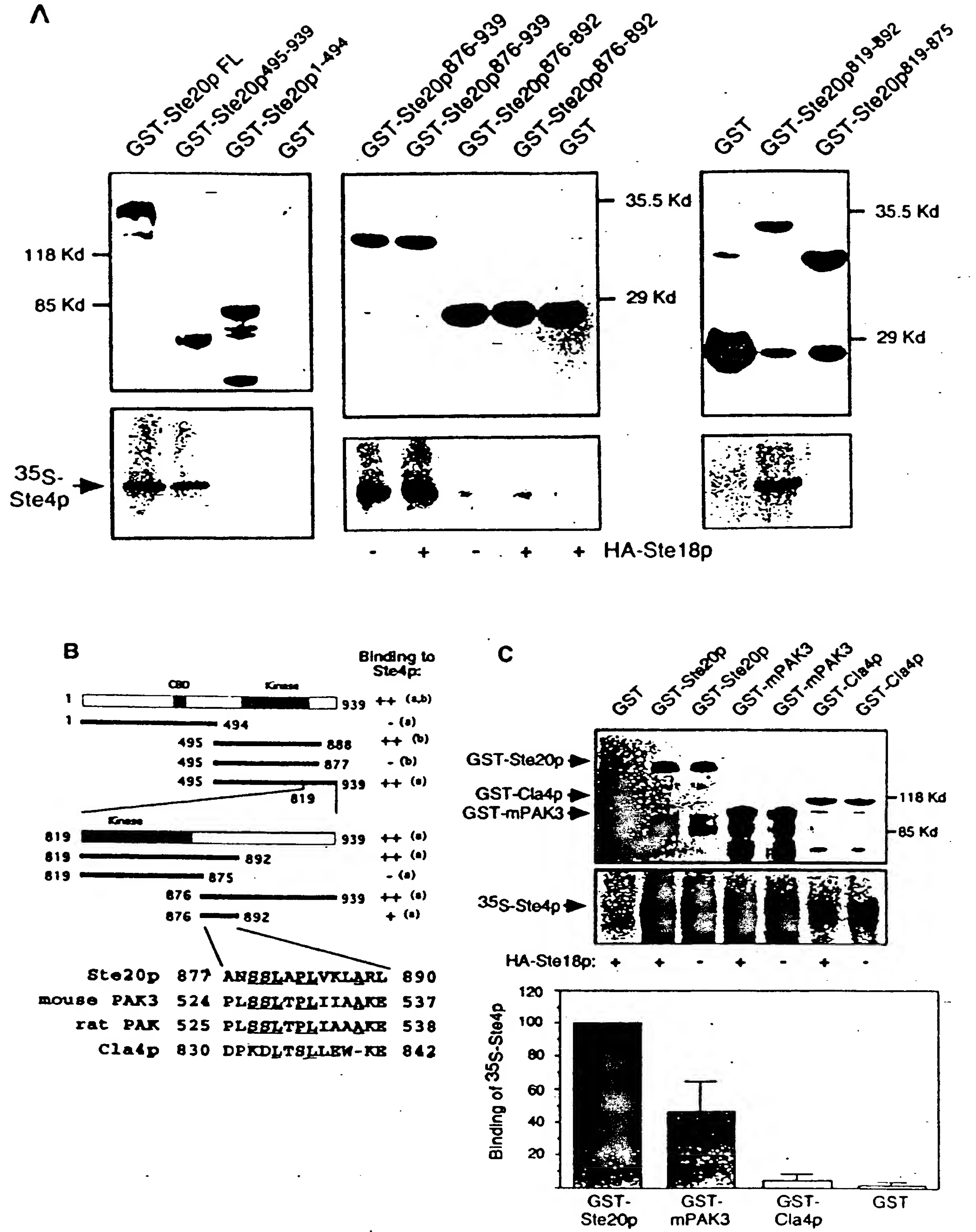
A
GST-Ste20p FL
GST-Ste20p495-939
GST-Ste20p1-494
GST
GST-Ste20p876-939
GST-Ste20p876-939
GST-Ste20p876-892
GST-Ste20p876-892
GST
GST-Ste20p819-892
GST-Ste20p819-875
118 Kd
85 Kd
35.5 Kd
29 Kd
35S-Ste4p
- + - + + HA-Ste18p
B
Binding to Ste4p:
CBD
Kinase
1 939 ++ (a,b)
1 494 - (a)
495 888 ++ (b)
495 877 - (b)
495 939 ++ (a)
819
819 939 ++ (a)
819 892 ++ (a)
819 875 - (a)
876 939 ++ (a)
876 892 + (a)
Ste20p 877 ANSSLAPLVKLARL 890
mouse PAK3 524 PLSSLTPLIIAAKE 537
rat PAK 525 PLSSLTPLIAAAKE 538
Cla4p 830 DPKDLTSLLEW-KE 842
C
GST
GST-Ste20p
GST-Ste20p
GST-mPAK3
GST-mPAK3
GST-Cla4p
GST-Cla4p
GST-Ste20p
GST-Cla4p
GST-mPAK3
118 Kd
85 Kd
35S-Ste4p
HA-Ste18p: + + - + - + -
Binding of 35S-Ste4p
120
100
80
60
40
20
GST-Ste20p
GST-mPAK3
GST-Cla4p
GST

FIGURE 2

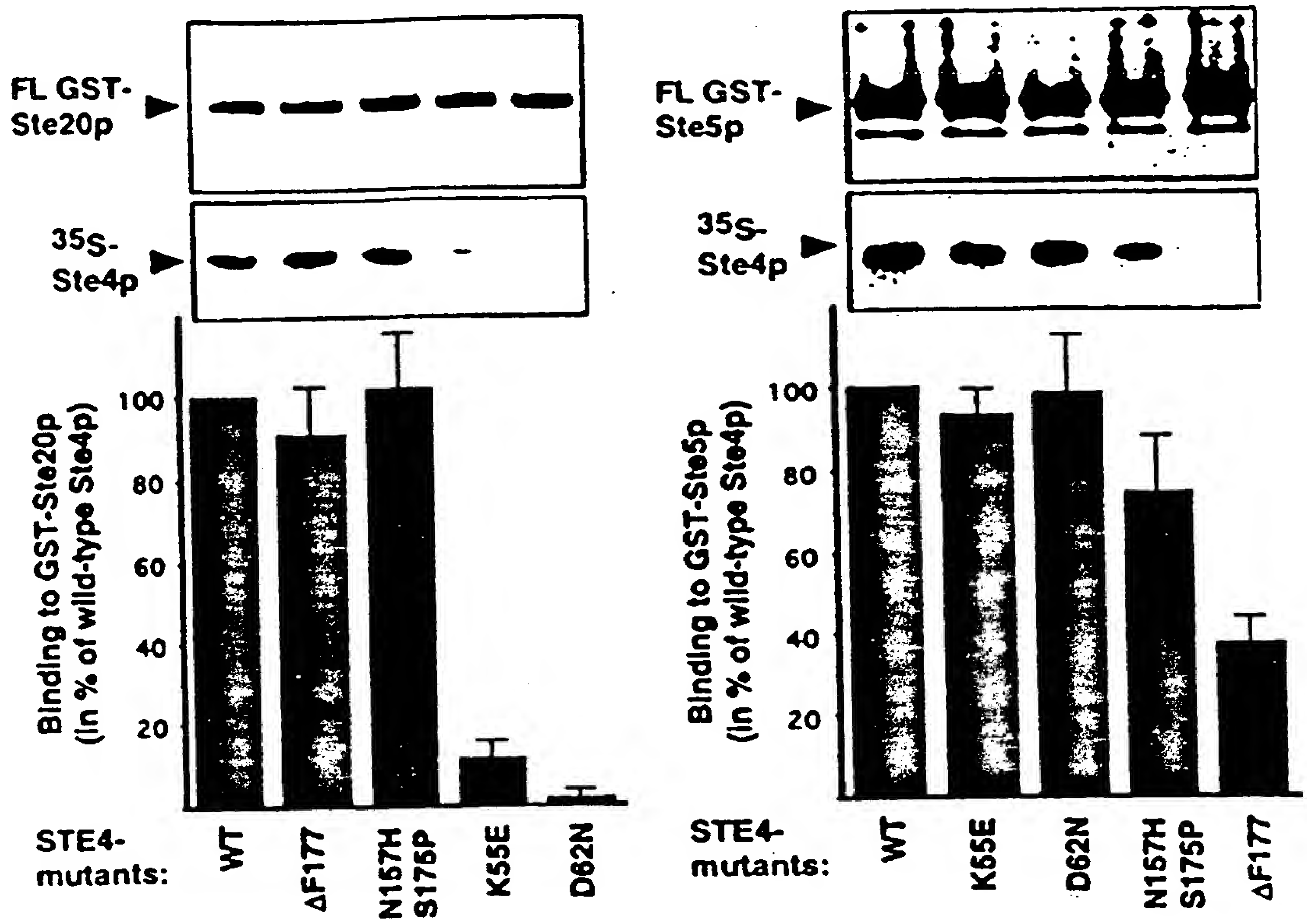
FL GST-
Ste20p
35S-
Ste4p
Binding to GST-Ste20p
(in % of wild-type Ste4p)
100
80
60
40
20
STE4-
mutants:
WT
ΔF177
N157H
S175P
K55E
D62N
FL GST-
Ste5p
35S-
Ste4p
Binding to GST-Ste5p
(in % of wild-type Ste4p)
100
80
60
40
20
STE4-
mutants:
WT
K55E
D62N
N157H
S175P
ΔF177

FIGURE 3

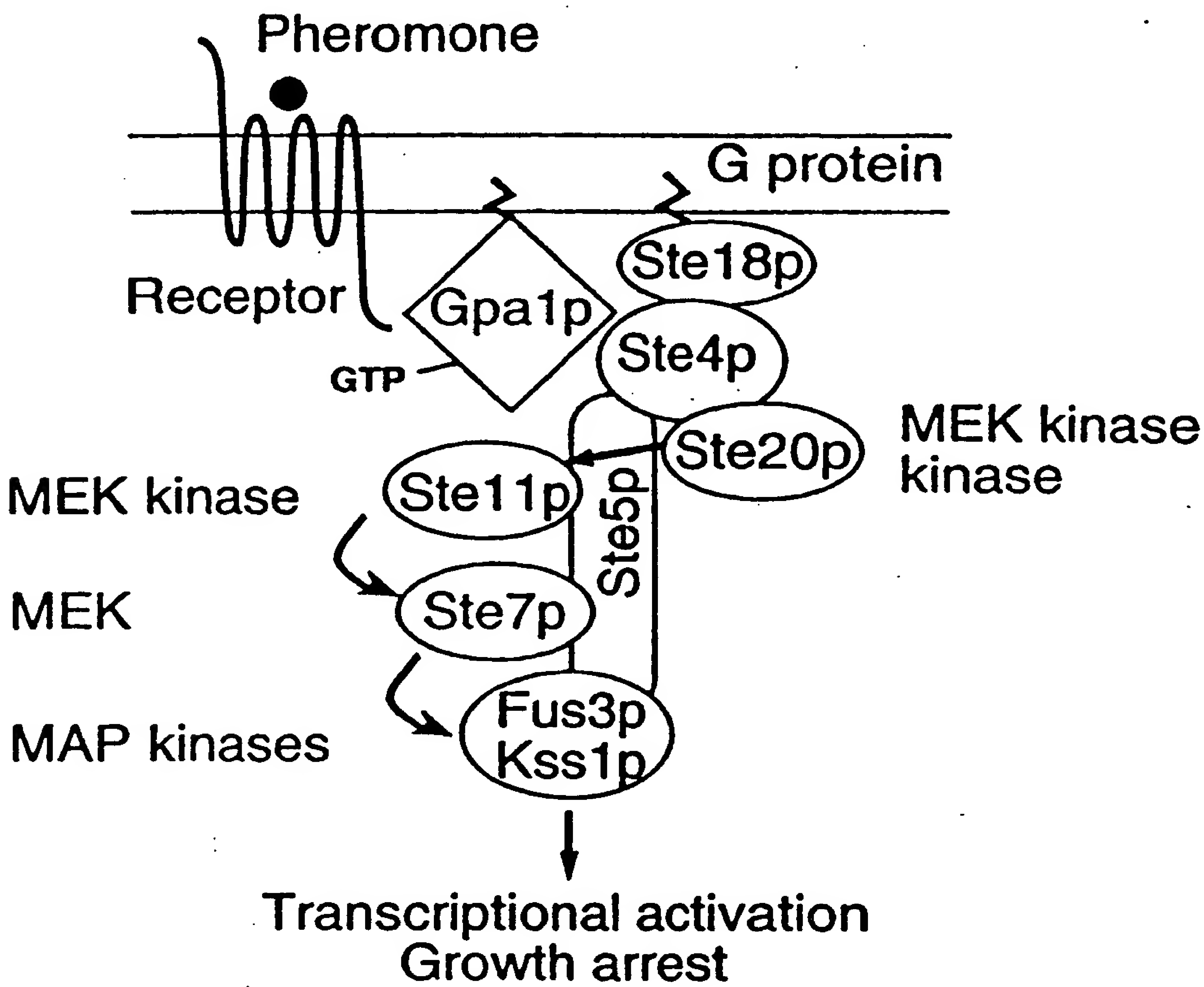
Pheromone
G protein
Receptor
Gpa1p
GTP
Ste18p
Ste4p
Ste20p
MEK kinase kinase
MEK kinase
Ste11p
Ste5p
MEK
Ste7p
MAP kinases
Fus3p
Kss1p
Transcriptional activation
Growth arrest

FIGURE 4

| | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Ste20p *(Sc)* | Q03497 | 876 | ean | **S** | **S** | **L** | A | **P** | **L** | **V** | K | **L** | **A** | **R** | lkkvaenmdad...//... | 939 |
| Cst20p *(Ca)* | Q92212 | 1209 | ddv | **S** | **S** | **L** | S | **P** | **L** | **V** | K | **I** | **A** | **R** | lkkmsesd | 1230 |
| Pak1/Shk1 *(Sp)* | P50527 | 641 | vpv | **S** | **S** | **L** | I | **P** | **L** | **I** | K | S | **I** | **H** | hsgk | 658 |
| Pak1 *(Hs)* | Q13153 | 525 | kpl | **S** | **S** | **L** | T | **P** | **L** | **I** | A | **A** | **A** | **K** | eatknnh | 545 |
| Pak2 *(Hs)* | Q13154 | 505 | kpl | **S** | **S** | **L** | T | **P** | **L** | **I** | M | **A** | **A** | **K** | eamksnr | 525 |
| Pak3 *(Hs)* | Q13177 | (473) | kpl | **S** | **S** | **L** | T | **P** | **L** | **I** | M | **A** | **A** | **K** | eamksnr | (493) |
| Pak1 *(Rat)* | P35465 | 524 | kpl | **S** | **S** | **L** | T | **P** | **L** | **I** | A | **A** | **A** | **K** | eatknnh | 544 |
| Pak2 *(Rat)* | Q62829 | 523 | kpl | **S** | **S** | **L** | T | **P** | **L** | **I** | L | **A** | **A** | **K** | eaiknssr | 544 |
| Pak3 *(Rat)* | Q64303 | 507 | kpl | **S** | **S** | **L** | T | **P** | **L** | **I** | L | **A** | **A** | **K** | eamksnr | 524 |
| Pak3 *(Rabbit)* | Q29502 | 504 | kpl | **S** | **S** | **L** | T | **P** | **L** | **I** | M | **A** | **A** | **K** | eamksnr | 524 |
| Pak3 *(Mouse)* | Q61036 | 523 | kpl | **S** | **S** | **L** | T | **P** | **L** | **I** | I | **A** | **A** | **K** | eaiknssr | 544 |
| DPak *(Dm)* | Q24190 | 685 | rpl | A | **S** | **L** | T | **P** | **L** | **I** | M | **A** | **A** | **K** | eatkgn | 704 |
| Pak1 *(Xen)* | (AF000239) | 504 | kpl | **S** | **S** | **L** | T | **P** | Y | **I** | I | T | **G** | **K** | qiakggh | 524 |
| Pak *(Ce)* | (D83215) | 547 | kpl | A | **S** | **L** | Y | Y | **L** | **I** | V | **A** | **A** | **K** | ksiaeasns | 569 |
| MIHCK *(Dd)* | (U67716) | 870 | cns | N | G | **L** | V | **P** | A | **I** | M | E | **A** | **K** | kakeahskfsih | 895 |
| MIHCK *(Ac)* | (U67056) | (279) | gpe | **S** | D | **L** | I | **P** | **L** | **V** | E | R | T | **K** | neaqrdfsmff | (303) |
| Cla4p *(Sc)* | P48562 | 829 | cdp | K | D | **L** | T | S | **L** | **L** | E | W | - | **K** | e | 842 |
| Cla4p *(Ca)* | (U87996) | 940 | gki | E | E | **L** | A | **P** | **L** | **L** | E | W | K | **K** | qqqkhqqhkqetsdtgfa | 971 |
| Skm1p *(Sc)* | Q12469 | 643 | csp | E | Q | **L** | K | V | S | **L** | K | W | H | | | 655 |
| Pak2/Shk2 *(Sp)* | (U45981) | 570 | cpt | E | D | **L** | K | S | I | **I** | F | S | R | **K** | anthin | 589 |

consensus **S S L** $\phi$ **P L** $^{I}{}_{V}$ x $\phi$ $\phi$ $\beta$

FIGURE 5

```
(P18851)  Ste4  maahqmdsitysnnvtqqyiqpqslqdisavedeiqnkieaarqeskqlhaqinkakhkiqdaslfqmankv
(P04901) Hgbb1                                     mseldqlrqeaeqlknqirdarkacadatlsqitnni
(P11016) Hgbb2                                     mseleqlrqeaeqlrnqirdarkacgdstltqitagl
(P16520) Hgbb3                                     mgemeqlrqeaeqlkkqiadarkacadvtlaelvsgl
(P29387) Mgbb4                                     mseleqlrqeaeqlrnqiqdarkacndatlvqitsnm
(P54314) Mgbb5                            matdglhenetlaslkseaeslkgkleeeraklhdvelhqvaerv
   Consensus    --------------------------------------------E---L------------D--L-------

(P18851)  Ste4  tsltknkinlkpnivlkghnnkisdfrwsrdskrilsasqdgfmliwdsasglkqnaipldsqwvlscaisp
(P04901) Hgbb1  dpv..griqmrtrrtlrghlakiyamhwgtdsrllvsasqdgkliiwdsyttnkvhaiplrsswvmtcayap
(P11016) Hgbb2  dpv..griqmrtrrtlrghlakiyamhwgtdsrllvsasqdgkliiwdsyttnkvhaiplrsswvmtcayap
(P16520) Hgbb3  evv..grvqmrtrrtlrghlakiyamhwatdskllvsasqdgklivwdsyttnkvhaiplrsswvmtcayap
(P29387) Mgbb4  dsv..griqmrtrrtlrghlakiyamhwgydsrllvsasqdgkliiwdsyttnkmhaiplrsswvmtcayap
(P54314) Mgbb5  eal..gqfvmktrrtlkghgnkvlcmdwckdkrrivsssqdgkvivwdsfttnkehavtmpctwvmacayap
   Consensus    ---------------L-GH--K-----W--D-----S-SQDG----WDS----K--A------WV--CA--P

(P18851)  Ste4  sstlvasaglnnnctiyrvskenrvaq.nvasifkghtcyisdieftd.nahiltasgdmtcalwdipkakr
(P04901) Hgbb1  sgnyvacggldnicsiynl..ktregnvrvsrelaghtgylsccrfld.dnqivtssgdttcalwdietgqq
(P11016) Hgbb2  sgnfvacggldnicsiysl..ktregnvrvsrelpghtgylsccrfld.dnqiitssgdttcalwdietgqq
(P16520) Hgbb3  sgnfvacggldnmcsiynl..ksregnvkvsrelsahtgylsccrfld.dnnivtssgdttcalwdietgqq
(P29387) Mgbb4  sgnyvacggldnicsiynl..ktregdvrvsrelaghtgylsccrfld.dgqiitssgdttcalwdietgqq
(P54314) Mgbb5  sgcaiacggldnkcsvypltfdknenmaakkksvamhtnylsacsftnsdmqiltasgdgtcalwdvesgql
   Consensus    S----A--GL-N-C--Y-------------------HT-Y-S---F------I-T-SGD-TCALWD------

(P18851)  Ste4  vreysdhlgdvlalaipeepnsenssntfascgsdgytyiwdsrspsavqsfyvndsdinalrffkdgmsiv
(P04901) Hgbb1  tttftghtgdvmslsl....apd..trlfvsgacdasaklwdvregmcrqtftghesdinaicffpngnafa
(P11016) Hgbb2  tvgfaghsgdvmslsl....apd..grtfvsgacdasiklwdvrdsmcrqtfighesdinavaffpngyaft
(P16520) Hgbb3  ktvfvghtgdcmslav....spd..fnlfisgacdasaklwdvregtcrqtftghesdinaicffpngeaic
(P29387) Mgbb4  tttftghsgdvmslsl....spd..lktfvsgacdassklwdirdgmcrqsftghisdinavsffpsgyafa
(P54314) Mgbb5  lqsfhghgadvlcldl....apsetgntfvsggcdkkamvwdmrsgqcvqafethesdvnsvryypsgdafa
   Consensus    ------H--D---L--------------F-S---D-----WD-R-----Q-F----SD-N-------G----

(P18851)  Ste4  agsdngainmydlrsdcsiatfslfrgyeertptptymaanmeyntaqspqtlkstssyldnqgvvsldfs
(P04901) Hgbb1  tgsddatcrlfdlradqelmtys................................hdnii..cgitsvsfs
(P11016) Hgbb2  tgsddatcrlfdlradqellmys................................hdnii..cgitsvafs
(P16520) Hgbb3  tgsddascrlfdlradqelicfs................................hesii..cgitsvafs
(P29387) Mgbb4  tgsddatcrlfdlradqelllys................................hdnii..cgitsvafs
(P54314) Mgbb5  sgsddatcrlydlradrevaiys................................kesii..fgassvdfs
   Consensus    -GSD-------DLR-D------S-----------------------------------------G--S--FS

(P18851)  Ste4  asgrlmyscytdigcvvwdvlkgeivgkleghggrvtgvrsspdglavctgswdstmkiwspgyq
(P04901) Hgbb1  ksgrlllagyddfncnvwdalkadragvlaghdnrvsclgvtddgmavatgswdsflkiwn
(P11016) Hgbb2  rsgrlllagyddfncniwdamkgdragvlaghdnrvsclgvtddgmavatgswdsflkiwn
(P16520) Hgbb3  lsgrllfagyddfncnvwdsmkservgilsghdnrvsclgvtadgmavatgswdsflkiwn
(P29387) Mgbb4  ksgrlllagyddfncsvwdalkggrsgvlaghdnrvsclgvtddgmavatgswdsflriwn
(P54314) Mgbb5  lsgrllfagyndytinvwdvlkgsrvsilfghenrvstlrvspdgtafcsgswdhtlrvwa
   Consensus    -SGRL----Y-D-----WD--K------L-GH--RV-------DG-A---GSWD-----W-
```

FIGURE 6